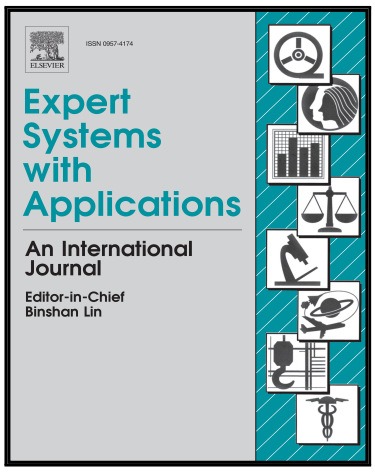
Journal Pre-proof



Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays

Marzia Hoque Tania , Khin T. Lwin , Antesar M. Shabut , Mohammad Najlah , Jeannette Chin , M.A. Hossain

PII: S0957-4174(19)30545-7
DOI: https://doi.org/10.1016/j.eswa.2019.112843
Reference: ESWA 112843

To appear in: *Expert Systems With Applications*

Received date: 13 November 2018
Revised date: 5 July 2019
Accepted date: 25 July 2019

Please cite this article as: Marzia Hoque Tania , Khin T. Lwin , Antesar M. Shabut , Mohammad Najlah , Jeannette Chin , M.A. Hossain , Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays, *Expert Systems With Applications* (2019), doi: https://doi.org/10.1016/j.eswa.2019.112843

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Ltd.

Highlights:

- An investigation into an automatic computational framework for lateral flow assays
- Critical examination of the data structure, features and algorithms
- Pseudo control colours were proposed as a new feature within the optimal feature set
- Proposed scheme offers high accuracy and fulfils ASSURED criteria

JOURNAL PRE-PROOF

Intelligent image-based colourimetric tests using machine learning framework for lateral flow assays

Marzia Hoque Tania[1], Khin T. Lwin[2], Antesar M. Shabut[3], Mohammad Najlah[4], Jeannette Chin[5], M. A. Hossain[2]

[1] Medical Technology Research Centre, Faculty of Science and Engineering & Faculty of Health, Education, Medicine and Social Care, Anglia Ruskin University, Chelmsford, UK

[2] School of Computing and Digital Technology, Teesside University, Middlesbrough, UK

[3] School of Arts and Communication, Leeds Trinity University, Leeds, UK

[4] School of Allied Health, Anglia Ruskin University, Chelmsford, UK

[5] School of Computing Sciences, University of East Anglia, Norwich, UK

Abstract

This paper aims to deliberately examine the scope of an intelligent colourimetric test that fulfils ASSURED criteria (Affordable, Sensitive, Specific, User-friendly, Rapid and robust, Equipment-free, and Deliverable) and demonstrate the claim as well. This paper presents an investigation into an intelligent image-based system to perform automatic paper-based colourimetric tests in real-time to provide a proof-of-concept for a dry-chemical based or microfluidic, stable and semi-quantitative assay using a larger dataset with diverse conditions. The universal pH indicator papers were utilised as a case study. Unlike the works done in the literature, this work performs multiclass colourimetric tests using histogram-based image processing and machine learning algorithm without any user intervention. The proposed image processing framework is based on colour channel separation, global thresholding, morphological operation and object detection. We have also deployed aserver-based convolutional neural network framework for image classification using inductive transfer learning on a mobile platform. The results obtained by both traditional machine learning and pre-trained model-based deep learning were critically analysed with the set evaluation criteria (ASSURED criteria). The features were optimised using univariate analysis and exploratory data analysis to improve the performance. The image processing algorithm showed >98% accuracy while the classification accuracy by Least Squares Support Vector Machine (LS-SVM) was 100%. On the other hand, the deep learning technique provided >86% accuracy, which could be further improved with a large amount of data. The k-fold cross-validated LS-SVM based final system, examined on different datasets, confirmed the robustness and reliability of the presented approach, which was further validated using statistical analysis. The understaffed and resource-limited healthcare system can benefit from such an easy-to-use technology to support remote aid workers, assist in elderly care and promote personalised healthcare by eliminating the subjectivity of interpretation.

JOURNAL PRE-PROOF

Keywords: Image processing; histogram thresholding; feature selection; machine learning; deep learning; colourimetric tests

1. Introduction

There is less than one physician per thousand population for more than 44% of the World Health Organisation (WHO) member states (World Health Organization, 2017). Even in a developed country such as the UK, there are only 2.806 doctors for every 1000 people. The longevity of human life has given rise to increasing the understanding of age-related disabilities and diseases, which can create significant burdens on already over-burdened healthcare systems. To support the elderly population, which is expected to increase to 1.91 billion in 2050 (P. D. of the Department of Economic and S. A. of the United Nations Secretariat, 2012), and limit the spread of pandemics, an intelligent, clearer and easier system with the least error-prone diagnosis results is required for both patients and clinicians. In the absence of expert clinical staff, there is a requirement for systems that are easily operable. Such a system could be used by aid workers in remote places to support primary healthcare, in time of epidemic and environmental monitoring for many purposes such as identifying safe drinking water. The easy-to-use system could support the growing need of the elderly population as well. Therefore, the early diagnosis facility, the disproportional ratio of health professionals (doctor, expert, staff, carer) to patients and the advancement of technology is influencing the field of healthcare prompting the sector of mobile phone-based microscopy, assays, and sensing platforms for Point-Of-Care (POC) diagnostics (Contreras-naranjo, Wei and Ozcan, 2016; Rajan and Glorikian, 2009).

The colourimetric tests are the prominent technologies used in the POC systems. The colourimetric tests for diagnosis purposes are being utilised for decades. The Lateral Flow Assay (LFA), a type of colourimetric test scheme, is more commonly either a qualitative or a semi-quantitative assay. The LFA are mainly popular for POC platforms since they are easy-to-use, fast and low-cost. However, they often suffer from limited specificity and sensitivity due to the limitation of materials including biochemical components (Koczula and Gallotta, 2016).

The naked-eye and colour chart based colourimetric tests including LFA expect the user to possess a perfect colour vision, whereas colour blindness is a common genetic deficiency. Globally, one in twelve men is colour blinded (NHS Choices, 2016). There are more than 2.7 million colour blind people in the United Kingdom (Colour Blind Awareness). Moreover, perception of colour can vary from person to person and reading from colour charts can be complicated for non-clinicians. As a POC system, the integration of computational system to

JOURNAL PRE-PROOF

LFA can enhance the overall diagnosis experience such as the research conducted by Ozkan (2017). In this work, we explored the computational solutions to provide an automatic colourimetric decision that fulfils our evaluation criteria.

This paper aims to provide a proof-of-concept for a dry-chemical based or microfluidic, stable and semi-quantitative assay using a larger dataset with diverse conditions. At first, this paper provides a context of the evaluation criteria and presents the rationale for the evaluation criteria (Section 2), followed by defining the assay types from the perspective of computer vision (Section 3). The current point of view in the field of POC systems is from a number of disciplines; dominated by biochemistry, nanotechnology and optoelectronics. Findings concerning such systems have presented the prospect of isolated individual colourimetric components but often lack the rigorous detailing of how such a system can be and should be designed. Therefore, there is a need for an extensive study to deal with the inadequacy to perceive colourimetric tests from the frame of computer vision. The challenges include the quest for searching a suitable image processing technique for robust operation of colourimetric tests. There is a requirement of knowledge exploration for such techniques to develop a better understanding of colourimetric test data, which can facilitate better management of the computational complexity of such data. Hence, experiments are designed accordingly; the detail of sample preparation is provided in Section 4, followed by the proposed image processing framework to separate the region of interest (ROI) from images of colourimetric tests.

Analysing the extracted features from ROIs could help to create domain-specific knowledge. Identifying the key features and how the features are being analysed can play a crucial role in the core model of a colourimetric Decision Support System (DSS). Thus, feature optimisation and feature analysis techniques would be a promising contribution, which is investigated in Section 6, and extended to exploring classification and regression algorithms and further expanded to consider the potential of utilising more advanced machine learning techniques in Section 7.

The purpose of this paper is to deliberately examine the scope of an intelligent colourimetric test that can reason about and interpret the colourimetric data, fulfils our defined evaluation criteria and demonstrate the claim as well. The paper also asserts the contribution regarding the pseudo-control colour.

2. Evaluation Criteria

WHO prefers the diagnostic system to be inexpensive, disposable and easy-to-use (Khademhosseini, 2011; Wang, Xu and Demirci, 2010). Such a diagnostic system should

JOURNAL PRE-PROOF

follow the criteria called ASSURED (Affordable, Sensitive, Specific, User-friendly, Rapid and robust, Equipment-free, and Deliverable) (Kettler, White and Hawkes, 2004). This paper studies, how a computational system can act as an expert to perform colourimetric test complying with the ASSURED criteria.

2.1 Affordable technology

The mobile phones have a high penetration rate (GSMA Intelligence), making it widely accessible and affordable technology to the resource-limited setting. By 2019, the number of mobile phone users is expected to reach 5.07 billion (Statista, 2015). From the computational context, the use of a mobile phone can act as an affordable-ASSURED technology. The mobile phone can effectively eliminate the operating cost by minimising the requirement of plate readers and analysers.

In general, paper-based assays such as Sicard et al. (2015) are more affordable and suitable for less trained personnel. A mobile-enabled paper-based assay can enhance the processing of result (Roda et al., 2016; Lopez-Ruiz et al., 2014), ease the effort to interpret the result and make the result conveniently communicable (Sicard et al., 2015). The objective of such a system is not aimed to replace the biochemical systems but instead to assist (Kim et al., 2017a), simplify (Lopez-Ruiz et al., 2014) or accelerate (Tania et al., 2017) the process. For example, when it is difficult to provide visually distinguishable colours, such systems can aid to provide a better decision (Abuhassan et al., 2017; Shabut et al., 2018). Therefore, this work aims to develop a system which is computationally efficient to be deployed on the mobile platform, making it an affordable system.

2.2 Specific and Sensitive Performance

The next ASSURED criteria are specificity and sensitivity, which require the system to have low false negative and false positives. Although, it is a common practice for the computation systems to present the result in terms of accuracy, evaluating the performance of the system only with accuracy can be misleading.

2.3 User-friendly System

The ASSURED criteria put emphasis on the minimum requirement of training from the users. The ratio between health professionals and patients are imbalanced worldwide. The global understaffed health systems can benefit from technologies that provide ease of use. These easy to use systems can support the associated need of growing elderly population, provide more autonomy to users for personalised healthcare at home settings, and more importantly in remote locations where there is a scarcity of trained medical personnel.

JOURNAL PRE-PROOF

Exploring the existing literature, this research suggests, the user-friendly system should not only require less medical training but also should not demand high technical skills from the user. Therefore, the system should require minimum user interactions or interventions with the system in order to provide a decision. In the reported articles on colourimetric tests, there are systems that require users to aid with the data pre-processing techniques e.g. cropping and seed points by the user (Rahmat et al., 2018; Mutlu et al., 2017; Solmaz et al., 2018). No detail description was provided in the article (Mutlu et al., 2017) regarding the cropping mechanism, therefore, the authors assume the cropping was performed manually without any intelligent image processing method. Solmaz et al. (2018) and Mutlu et al. (2017) utilised smartphones, whereas Rahmat et al. (2018) used a scanner for image acquisition. The cropping techniques used in these articles eliminated the segmentation process. However, the technique compromises the ease of use, compelling the user to possess technical skills.

2.4 Rapid and Robust System

The next criteria are to make the system rapid and robust. This work mainly focuses on machine learning based computational systems. To evaluate the rapidness, the training time should be taken into account so that the system can support in time of any sudden outbreak such as climate change-related diseases (Kabir, Rahman and Milton, 2014). However, the testing time should be considered as the execution time. On the other hand, robustness is a vast term and has a direct impact on the ease of use of the system. In order to provide reproducible and accurate results, if there are many rigid guidelines for the user to follow, then the system preserves lower robustness as well as reduced user-friendliness. In this case, the robustness could be in terms of the format of the data or regarding the environmental settings.

The lighting condition is one of the biggest concerns for image processing, especially medical imaging. The ambient light can have a huge impact on the RGB value. Therefore, lighting condition is considered as a key factor of the robustness (Solmaz, 2018; H. Kim, Awofeso, Choi, Jung, & Bae, 2017; Mutlu et al., 2017). Solmaz et al. (2018) considered 7 illumination conditions to train the model. The light condition can be channelled through additional optomechanical attachments such as H. Kim et al., 2017. The algorithm proposed by Mutlu et al. (2017) for the classification of pH test strips showed equal performance with and without such hardware attachments. The study (Mutlu et al., 2017) considered three different lighting conditions while experimenting without any hardware.

The camera to sample position can influence the image processing requirements as well as the classification performance. The mobile applications available in the app stores utilise a guideline or virtual plate to limit the location of the sample position as well as the distance of

JOURNAL PRE-PROOF

the camera (Sicasys Software GmbH, 2017; Enzo Life Sciences inc., 2015; Alidans srl, 2015). Mutlu et al. (2017) utilised six different orientations of the pH test strips (Merck, Germany), effectively varying the position of the sample from the camera lens as well as from the source of the light. Although the purpose of the variety in the rotation was to train the system for robust orientation of the sample, the images were manually rotated before training to maintain the same alignment. However, the system does not include any automatic image processing technique, rather rely on the user to crop the image.

Another parameter for robustness is interoperability (e.g. hardware compatibility, application programming interface or API). A mobile-enabled algorithm was tested on the paper strips to present a system with inter-phone repeatability by Yetisen, Martinez-Hurtado, Garcia-Melendrez, da Cruz Vasconcellos, & Lowe (2014). The downside of the system is longer calibration time. The performance of another smartphone application for paper-based saliva-alcohol testing was evaluated by (Kim et al., 2017a) on five different smartphones, effectively varying the hardware and software components including phone-camera. However, the system involves additional hardware for illumination and imaging consistency.

Therefore, the criteria of robustness might be immeasurable, because how much autonomy a single system can provide without compromising the rest of the parameters of the ASSURED criteria is an optimisation issue. In this work, we evaluate the robustness or R-criteria by computational complexity and adaptability.

2.5 Equipment-free System

The next ASSURED criterion is whether the system equipment free. This criterion is closely associated with the cost of the system. If the system can turn the smart device such as mobile phone and tablet into colourimetric test reader, essentially there is no need for additional equipment. The standard practice for colourimetric tests involves plate reader e.g. 96 well plate reader utilising light absorbance for wet-chemical based quantitative tests, analysers for paper-based test strips and naked-eye tests for qualitative tests. In literature, substantial amount of the reported articles on mobile-enabled systems utilised additional hardware attachment to channel the lights, enhance the image or ease the image processing technique (Kim et al., 2017a; Hussain et al., 2017; Kim et al., 2017b; Masawat, Harfield and Namwong, 2015).

2.6 Deliverable System

The last criterion is- the system should be deliverable implying it should be accessible by those who will use the system. There is a close relation between affordable and deliverable systems. According to the World Bank, more households are likely to possess mobile phones

JOURNAL PRE-PROOF

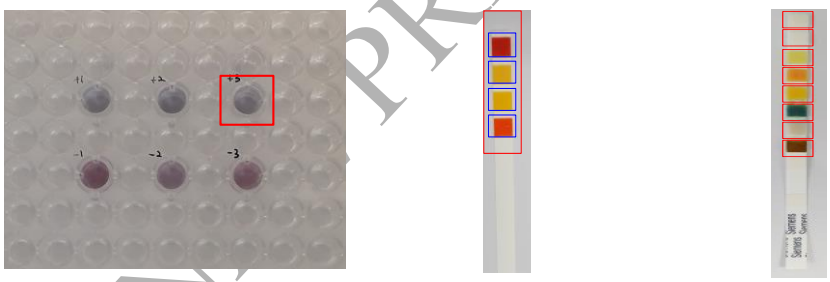
than a toilet (World Bank Group, 2016). Therefore, a computationally efficient system to be deployable as a mobile-enabled system is supposed to be more accessible than the specialised devices.

Therefore, the aim of this work is to design and develop a system that fulfils ASSURED criteria. Our goal includes a system that is intelligent to make it automatic and user intervention free; computationally efficient to be deployed in the mobile environment; adaptable to comply with variability; computationally efficient, making the system rapid and the specificity and sensitivity should be balance and high to produce a reliable accuracy.

3. Assay Type

The assay type can be defined in many ways such as based on time point, number of analytes, signal amplification method, type of the substrate and format of the result. In terms of detection method, this work considers the assay involving visible colour spectrum only and based on the format of the result, this paper mainly focuses on the semi-quantitative assay.

Based on the number of analytes, the colourimetric assay can be a single target or multiplex. However, this work considers sample in image format. Therefore, it is important for the decision making elements to realise the number of objects associated with a single sample.



(a) (b) (c)

Fig. 1: Number of targets in the image. (a) TB-test: image contains six filled wells. In other words, there are six samples. From the computer vision perspective, there are six objects. (b) pH indicator paper: four objects separated by blue boxes belong to the same pH test, (c) Multiplex assay: each object is directed to a different target

i) Single object/ Sample:

Different type of targets and objects associated with the colourimetric test are shown in Fig. 1 for the readers to visualise the variety of the ROI within the scope of this research area. In Fig. 1(a), the ROI is the samples or filled wells in a 96-well assay plate involving plasmonic enzyme-linked immunosorbent assay or ELISA (Shabut et al., 2018). There are six samples in the image. A single sample is outlined with a red box. In this case, the

single object or well represents a single sample obtained from a single source or subject i.e. urine of an individual and a single test i.e. TB-antigen specific antibody detection. For an image of 96-well plate, there can be multiple samples (maximum 96 samples). Our earlier works involve such wet-chemicals based colourimetric test (Tania, Lwin and Hossain, 2016; Abuhassan et al., 2017; Shabut et al., 2018; Tania et al., 2018).

ii) Multiple objects/Sample:

The main focus of this paper is the type of assay shown in Fig. 1 (b), which is an image of a universal pH indicator paper. The ROI is the colour pads. Depending on the brand, the commercially available pH indicator paper can have one or multiple colour pads to provide a decision on a single pH level. The pH paper strips using multiple colour pads for a single test type are more reliable. Therefore, this research involves such a test paper for pH level detection.

Table 1: Region of interest, feature-set and classification algorithm for paper-based assays

Reference	Object/ sample	Histogram Features	Colour Spaces	Algorithm	Result Accuracy
H. Kim et al. (2017)	Single	Mean, Median, Mode, Bin median and Bin centroid	RGB, HSV, YUV, LAB	LDA, SVM, ANN	80%; 100%
Solmaz et al. (2018)	Single	With and without grey-world corrected mean	RGB, HSV, LAB	LS-SVM, RF	90.3%; 95%
Mutlu et al. (2017)	Multi	Mean of JPEG, RAW and RAWc	RGB	LS-SVM	100%
Rahmat et al. (2018)	Multi*	Mean	RGB	Euclidean distance	95.45%

*Multi-object per sample but single object per target (Multiplex assay)
LDA: Linear Discriminant Analysis; SVM: Support Vector Machine; ANN: Artificial Neural Network; LS-SVM: Least squares support vector machine; RF: Random Forest

The reported articles using paper-based assays for the various application considered both single colour pad or object per sample as well as multiple ROIs or objects that belong to the same test (Table 1). In this paper, multiple objects i.e. colour pads belong to the single sample as well as a single test. In Fig. 1 (b), the image of a paper strip indicating pH 3 has 4 colour pads. This paper mentions the block of colour pads as objects/ sample. These objects are collectively responsible to produce a decision. As these objects, outlined in blue boxes,

are detected individually in this work, the features per colour pad are multiplied by the number of objects/ sample.

iii) Multi-test/ Image and Single object/ test:

The multiplex assays are capable to perform multiple tests on a single sample using the same test strip. Each object, block or colour pad in Fig. 1(c) represents different test type such as glucose, pH, protein and ketone. In this paper, we have used an intelligent image processing scheme to detect objects/sample in the pH test strip. As the objects are detected individually, the same approach can be implemented to the multiplex assays, e.g. urine dipstick test as well. Therefore, in addition to the pH test strips, for the proof of concept, a separate dataset of untested blank urine dipstick was utilised to evaluate the strength of the image processing algorithm, which is one of the contributions of this paper.

4. Experimental Selection

4.1 Materials Preparation

As a paper-based LFA, this work mainly focuses on pH indicator universal test strips. The pH ranging from 3.0 to 10.0 was considered for this work. As the objective of this work is to deliver a proof of concept, the whole range (0-14) was not considered. In this work, we used buffer solutions to ensure more stability and longevity of the solution, therefore increasing the reliability of the overall experiment.

For pH 3.0-5.0, the citrate buffer was prepared as x ml $A + y$ ml B diluted in deionised water, where A= X citric acid ($C_6H_8O_7$) and B= Y sodium citrate ($Na_3C_6H_5O_7$), where X and Y represents the concentration in molar or M (Gomori Buffers). Traditionally 0.1M is used. However, a higher concentration would result in longevity of the solution. Similarly, for pH 6.0-7.0, the phosphate buffer was prepared from NaH_2PO_4 and Na_2HPO_4. Using $NaHCO_3$ as the weak base and Na_2CO_3 as the strong base, buffer solution for pH 9 was prepared. The pH 8 and 10 were tested by commercially available NIST traceable borate pH buffer solution (Fisher Scientific, UK). The pH level measurements were controlled with calibrated pH meter (HI 208, Hanna Instruments). The calibration was conducted with standard buffer of pH 4 and pH 7.

The Fisherbrand® pH test strips were immersed in the prepared pH solutions. The test strips instantly form the colour and change it quickly after drying. Therefore, images were captured rapidly. Some of the test strips were allowed to dry on tissue paper, while some of them were purposefully not allowed to dry the residue on tissue paper. The purpose of it was to make the dataset more robust.

JOURNAL PRE-PROOF

4.2 Experimental Setup and Dataset

In this work, the images were captured by Samsung Galaxy S6. The SM-G920F camera (f/1.9, exposure time 1/50 second, focal length 4mm, maximum aperture 1.85, 35mm focal length =28, normal exposure program, without flash) was set on its default mode. The idea is to utilise the strength of the available camera with minimum user interactions. The camera automatically adjusted the white balance, ISO speed, metering mode and set the brightness by the internal software of the smartphone. The detail of the experimental setup is provided in the Supplementary Document 1.

4.3 Dataset

The original dataset in Table 2 is utilised to investigate the required image processing framework, classification algorithm and an optimum number of features. The original dataset contains 520 images. Let us denote the samples of pH 3-9 in the original dataset as 'D-o1-9'. Once the model was developed, the rest of the datasets were utilised to explore the merits and capabilities of the proposed scheme.

Table 2: Dataset

Sl.	Name	Description	Brand	Lighting Environment	Device	Samples/ class	Total
1	original	Universal pH indicator paper	Fisherbrand® pH	$\tilde{A}_e = A_i + A_c(l)$	Samsung Galaxy S6	65	520
2	new_pH10			$\tilde{A}_e = A_i + A_c(l)$	Samsung Galaxy S6	65	65
3	D-A_i			Images captured in without any controlled light	Samsung Galaxy S6	15	120
4	D-dl			$\tilde{A}_e = A_i + A_c(dl)$	Samsung Galaxy S6	15	120
5	D-jl			$\tilde{A}_e = A_i + A_c(jl)$	Samsung Galaxy S6	15	120
6	D-i			$\tilde{A}_e = A_i + A_c(l)$	iPad Pro	20	20
7	D-j			$\tilde{A}_e = A_i + A_c(l)$	Samsung	20	20

JOURNAL PRE-PROOF

					Galaxy J3 Prime		
8	ℵ_pH		Hicarer-pH Test Strips-01	$\tilde{A}_e = A_i + A_c(l)$	Samsung Galaxy S6	65	65
9	untested urine dipstick	Reagent strips for urinalysis	Multistix® GP, Siemens	$\tilde{A}_e = A_i + A_c(l)$	Samsung Galaxy S6	5	5
Total							**1,055**

$$D\text{-}A_i + D\text{-}dl + D\text{-}\,jl = D\text{-}lights$$

In literature, Mutlu et al. (2017) used single test strip per class and extended the dataset by changing the format of the file, the orientation of the test strip and capturing the image of the same strip for 5 times. Their study did not repeat the test on pH strips to consider potential anomalies. They tried to compensate for the issue by capturing the image of the same test strip using different controlled-lighting conditions. The authors acknowledge that presented case study is a stable assay which is not subject to as many variations as Shabut et al. (2018). However, the dataset of Dhar, Mehta, & Sit (2017) indicated that even a stable paper assay can have nonconformities when the solution is not held within the colour pads and it is spread across the base paper, which can increase the false positive ROI area during image processing and act as a noise during classification. Consideration of such noise level at different test attempt would have asserted statistical likelihood and demonstrated a more reliable system by Mutlu et al. (2017). Therefore, in order to rectify this issue, in our paper, the dataset contains 65 images of individual pH indicator strips for each of the mentioned levels. The samples were allowed to have any arbitrary orientation as well as position within the sample plane exposed to the camera (Fig. S1.2).

JOURNAL PRE-PROOF

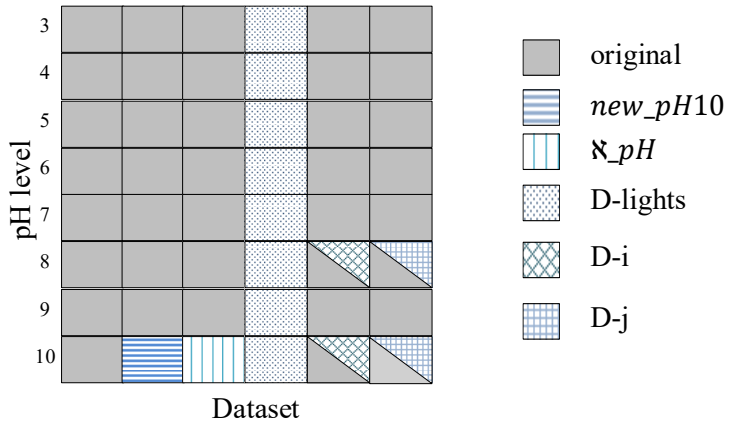


Fig. 2: Proportion and distribution of the dataset

The original dataset was utilised to develop the system from image pre-processing to feature extraction and classification. Using 10-fold cross-validation, the dataset assisted in optimising the features and identifying the best suitable classifier. Later, 535 images were considered as the extended dataset to evaluate the concept using a wider latitude. The proportion of data distribution is illustrated with a heat-map (Fig. 2).

The purpose of using these extended datasets was to evaluate the robustness of the image processing algorithm, the stability of the classifier and reliability of the overall system by investigating the following questions.

i) How the system performs on an unseen data under a similar condition?
   - (new_pH10)
ii) Can the system be trained for a similar assay from a different commercial brand?
   - ($\aleph\_pH$)
iii) Does the image process algorithm adaptable to analogous assay type?
   - (untested urine dipstick)
iv) Is the system trainable for images under different illumination condition without fine-tuning the image processing framework, feature set and hyper-parameters? -(D-lights)
v) Is the system trainable for images captured by different devices without fine-tuning the image processing framework, feature set and hyper-parameters?
   - (D-i and D-j)

Extended Dataset

Among these extended samples, aside from the urine dipstick, the rest of the samples were immersed in pH buffer solutions.

JOURNAL PRE-PROOF

i) At first, a dataset (new_pH10) was created using the same assay brand and same ambient condition, placing the test strips in an arbitrary position within the sample plane with the aim to create more diverse conditions. This new_pH10 dataset was used as a testing dataset to validate the system on the entirely unseen data.

ii) In order to validate the robustness of the image processing algorithm, universal pH indicator paper of a different brand (Hicarer-pH Test Strips-01) was utilised, denoted as $א\_pH$ papers. The dimension, including the thickness of the colour pads of $א\_pH$ papers is slightly different than the original dataset. The colour pads in $א\_pH$ papers are densely situated and the base papers are more hydrophilic than our original test strips. Therefore, the image-processing technique would have to deal with more noise for these test papers.

iii) In the reported articles (Yetisen et al., 2014; Rahmat et al., 2018; Smith et al., 2016; Chen, Wu and Dong, 2014; Wirth et al., 2018), the urine dipstick is a well-utilised example of LFA using with and without additional hardware. Due to the resemblance of the assay, a urine dipstick was included in the extended dataset. Similar to these multi-object assays, e.g. original dataset, the urine dipstick has multiple colour pads. Although these assays are different in terms of the number of targets, the image processing framework should be able to segment the ROIs. Image of a single untested urine dipstick was captured five times with a slight variation in the position for repeatability. Due to different targets, the dataset was not used for classification.

iv) The reported articles emphasize creating a diverse dataset by considering different illumination conditions and mobile devices (Kim et al., 2017a; Solmaz et al., 2018). Therefore, we have further extended our dataset (Table 2) to include the additional three different illumination conditions as described in the experimental setup. The original dataset was generated in an indoor laboratory environment. Without the ceiling lights (in the presence of A_i), for each pH level, 15 samples were generated- a) without any controlled light, using natural daylight only (D-A_i) b) using warm light (D- jl) and c) using cold light (D-dl). Therefore, from 120 independent pH tests, 360 images were produced using these lighting conditions (D-A_i, D- jl, D-dl). From this point forward, these images are collectively denoted as 'D-lights'.

v) The properties of images captured with different devices may vary due to the camera, optics and imaging software (Solmaz et al., 2018), even when the interoperability issue (Yetisen et al., 2014) of the mobile application is not considered. Hence, a small

JOURNAL PRE-PROOF

dataset[1] (D-i and D-j) was generated from images captured using different devices to explore the impact (Table 2).

To summarise the overall dataset, this work includes an LFA dataset of 1,055 images in total.

This paper investigated the colourimetric classification method using both traditional machine learning as well as deep learning techniques. The deep learning technique is capable to circumvent the effort to find the perfect features for an image-based system. However, the traditional machine learning techniques, both classification and regression, require image processing and feature extraction to provide the features as the input of the classifier or regressor. Therefore, this paper sequentially discusses the data pre-processing steps for classification and regression, followed by a separate attempt using deep learning models to provide a colourimetric decision.

5. Image processing of the paper strips

An expert system is a knowledge-based system that employs knowledge about its application domain and uses an inferencing (i.e. reason) procedure to solve problems that would otherwise require human competence or expertise. The arrangement expert systems in the imaging context have been utilised in diversified fields as exemplified in (Janke, Castelli and Popovič, 2019; Seo and Shin, 2019; Dang et al., 2019; Shabut et al., 2018; Carbonera, Abel and Scherer, 2015). This paper utilises such an expert system that can interpret the visual knowledge to identify objects in the assay based on the stated characteristics and provide a classification of the colourimetric test. The use of an intelligent image processing scheme can make a colourimetric system more user-friendly[2] by reducing the required user intervention to produce the result. A robust image processing algorithm can also eliminate the necessity of the additional hardware attachments.

[1] The dataset of D-i and D-j are relatively small (<30 samples for 2 classes). Rest of the pH samples per class, e.g. original dataset, contain 65 images per class. Therefore, D-i and D-j

[2] The authors are not referring to user interface. The user-friendly system is defined in the introduction, i.e. a system which promotes ease of use and requires less medical or technical trainings.

JOURNAL PRE-PROOF

Table 3: Variability regarding imaging

Premise	Parameter	Specifics
Camera parameter	Resolution (MP)	Low: 3.2 (Cooper et al., 2012) High: 20.7 (Kim et al., 2017a)
	ISO	Varying from 50 (Alankus et al., 2018) to 800 (Karlsen 2018; Lopez-Ruiz et al., 2014). Auto (Karlsen 2018)
	Other parameters	Constant (Mutlu et al., 2017) Auto (Solmaz et al., 2018)
Camera to sample position	Distance	Low: 5cm (Yetisen et al., 2014) High: 2 feet (Feng et al., 2014)
	Exposure (Angle)	Parallel (Alankus et al., 2018) Tilted (Karlsen and Dong 2017)

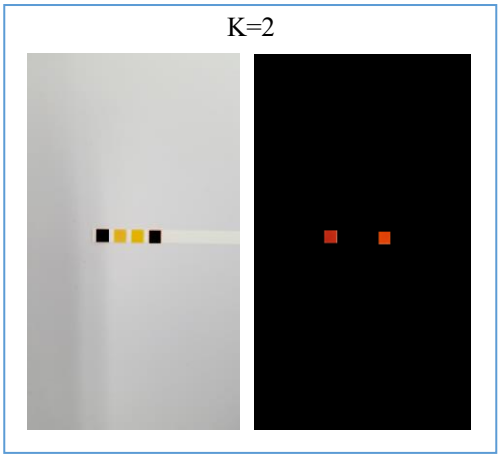
Image acquisition is an essential step in an image-based system. The imaging condition can undoubtedly influence the image quality as well as colours. A wide range of smartphone camera specifications along with intrinsic and extrinsic parameters has been explored in the literature such as the International Organization for Standardization (ISO), flash and focus. Such explorations are exemplified in Table 3. Within this context, an important parameter to be considered is the distance between the camera and the sample. Alankus et al. (2018) kept this distance as 30 cm and Solmaz et al. (2018) as 16cm, whereas Yetisen et al. (2014) chose the distance to be rather small by setting it as 5 cm. Sicard et al. (2015) defined close distance as 6cm away. Feng et al. (2014) kept the Google Glass more than 2 feet away to capture good quality images. Our research suggests that instead of setting a hard constraints on the distance between camera and sample, a study should consider the following while capturing images: i) the primary objective should be a good quality image; ii) sample plane should be exposed to the camera in a certain way that ROI is clearly visible, without any blurriness, hence focus should be adequate, which can be easily handled by modern phone cameras and iii) capability of the image processing algorithm. The image processing algorithm may use object detection based on geometric features. In such cases, the physical size of the sample in an image may play a decisive role. Advanced image processing algorithm such as deep learning is known for its ability to recognise objects despite its size. However, a heavy algorithm such as deep learning might not be suitable for real-time native application in resource-limited settings. Another circumstance regarding this distance issue could be the noise filtering process in an

JOURNAL PRE-PROOF

image processing algorithm. The conventional noise filtering, for example, the morphological operation may define the noise by connectivity. Therefore, a guideline should be provided to the user, if the system gets easily affected by any variation in this distance. On the other hand, a rigid guideline would affect ease-of-use. Under the circumstances, this paper maintains the camera to sample position within close proximity as most of the reported articles (Yetisen et al., 2014; Shabut et al., 2018), retaining the average 35mm focal length around 28 as mentioned in Section 4.2.

In our earlier work (Abuhassan et al., 2017; Shabut et al., 2018) on wet-chemicals based colourimetric test, we segmented the images using k-means clustering (Arthur and Vassilvitskii, 2007), which was an iterative method that was responsible to generate an enormous amount of garbage-files. As it can be visualised from Fig. 1(a), the colour of the samples of the qualitative test are bluish pink and pinkish blue. The R and B channels had an overlap. Therefore, k-means provided a better performance among the histogram thresholding based image processing techniques. In this work, the associated colour of the samples are mostly colour opponents. Using k-means would create more garbage-files to accommodate the images of multiple clusters before selecting the best cluster. Moreover, due to multiple colour opponent objects of the same sample, it would be difficult to keep these colour pads in the same cluster. This issue can be visualised in Fig. 3.

JOURNAL PRE-PROOF



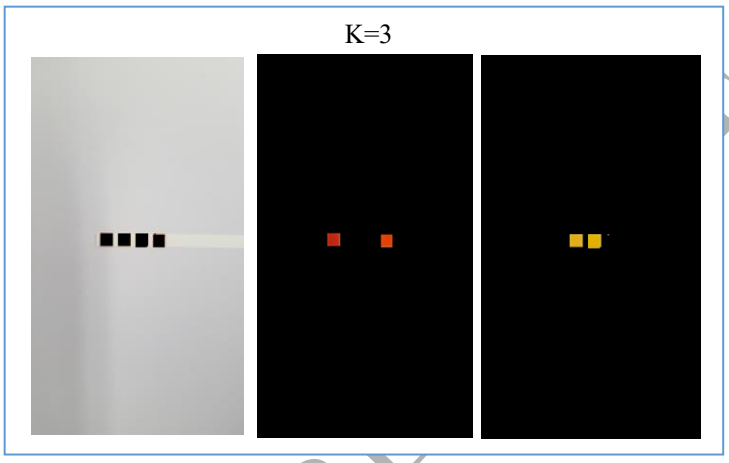


Fig. 3: kmeans clustering (Arthur and Vassilvitskii, 2007), varying k=2 to 3.

Being a 'multi-object/ sample' type colourimetric test, the ROIs divided into different clusters would occupy more memory space and the feature extraction stage would require more iterations, essentially taking more time to process. Therefore, exploring several image processing algorithms as shown in Fig. 4, the context is provided in (Shabut et al., 2018; Achanta et al., 2012; Bradley and Roth, 2007; Otsu, 1979), this paper presents an image processing framework circumventing the iterative approach, making it more suitable for mobile environments of limited storage and processing capacity.

JOURNAL PRE-PROOF

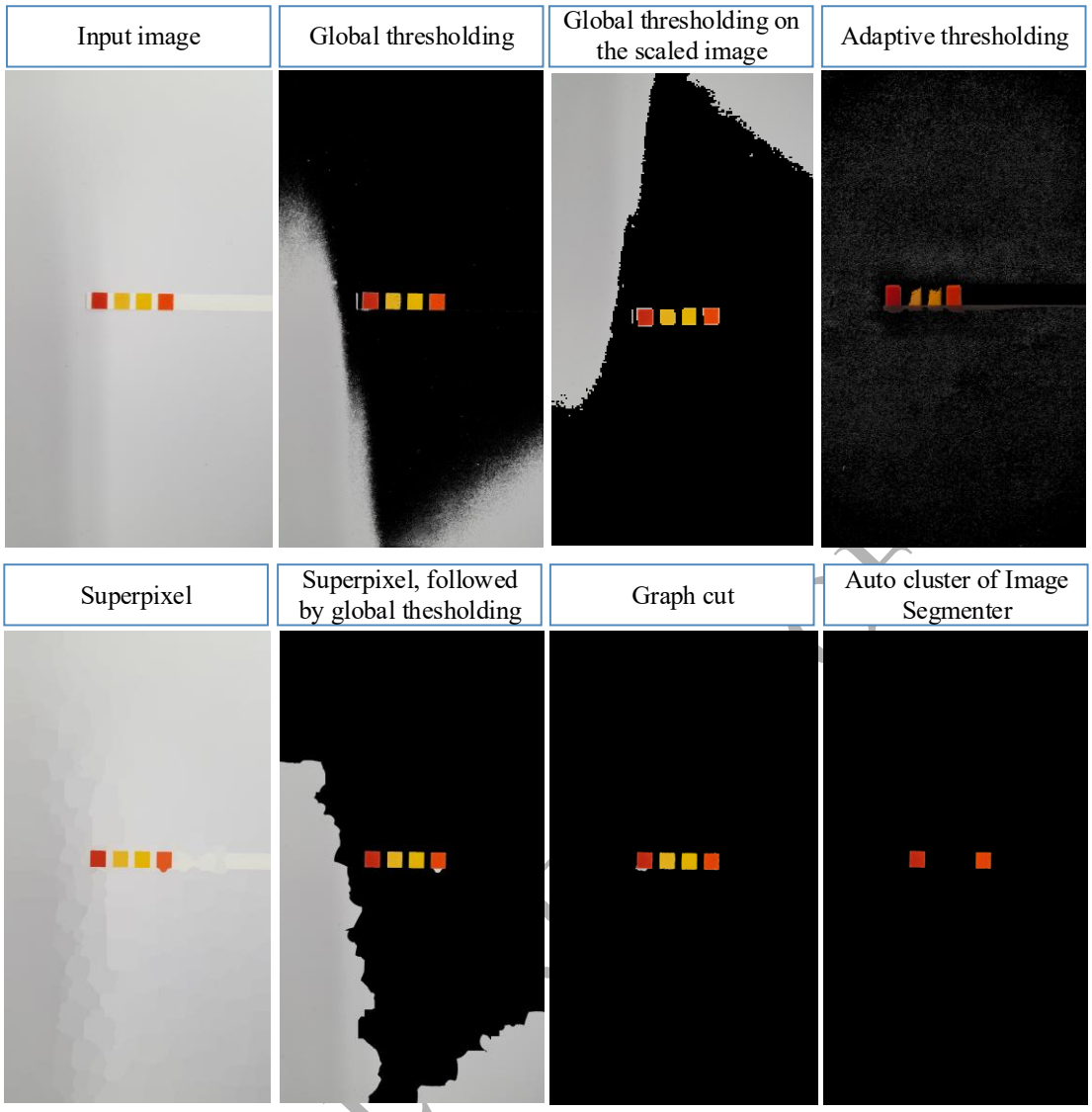


Fig. 4: Exploration of relevant image processing algorithms exemplified using an image of class label pH 3

The key steps of the algorithm are illustrated as a framework in Fig. 5. The purpose of this detailed framework was to ease the consequent processing and reproducibility of the outcome.

JOURNAL PRE-PROOF

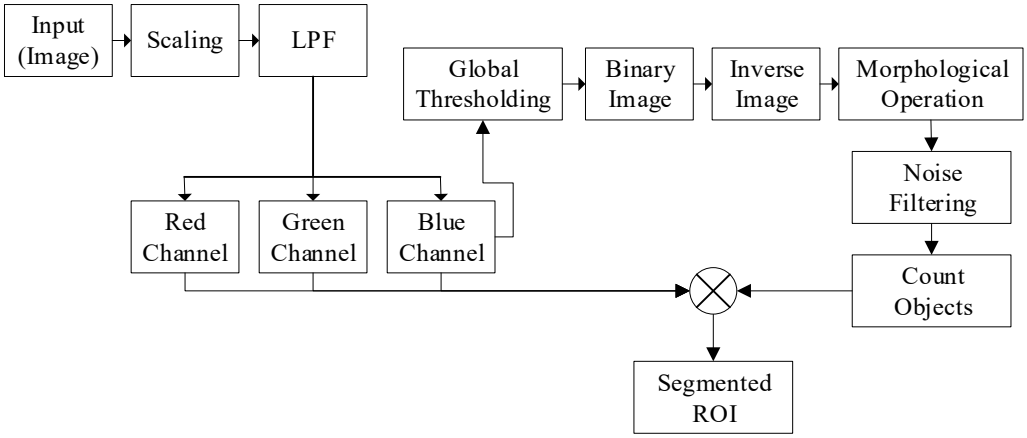
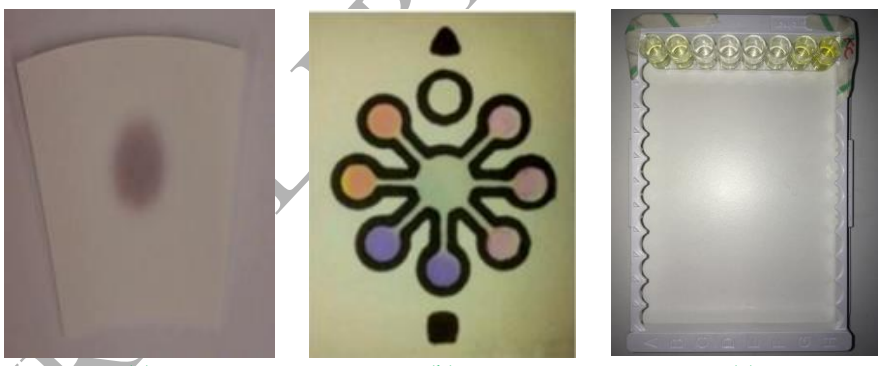


Fig. 5: Image processing framework

At first, the dimension of the images was reduced by dynamic scaling (Supplementary Document 2). For a known imaging condition, the height and the width of the image will not vary to a great extent. However, it may vary due to the factors such as the position of the camera, size of the plate and camera configuration (Fig. 6). Thus, the size reduction was performed dynamically based on the size of the input image and proportionally so that the geometry of the ROI was not deformed, giving the users more flexibility towards imaging condition.



(a) (b) (c)

Fig. 6: Variation in the size, shape and location of the objects. (a) Reaction kinetics of alkaline phosphatase on paper[3], where the outline of the shape of the object is ambiguous (b) pH and nitrite detection by (Lopez-Ruiz et al., 2014), which is case of multi-object and multi-test per image, (c) ELISA for analysing pro-inflammatory protein production (IL-6)[4], where

[3] Courtesy: Dr Mohidus Samad Khan, Department of Chemical Engineering, Bangladesh University of Engineering and Technology, Bangladesh

[4] Courtesy: Charys Palmer and the Department of Biomedical and Forensic Science, Anglia Ruskin University.

JOURNAL PRE-PROOF

the position of the samples and closely associated background and sample colour is making the object detection challenging

Following size reduction, colour filtering using a low pass filter (LPF) was applied to allow only a specific spectrum from the RGB values.

x_r = 8-bit unsigned integer array of filtered image in red channel

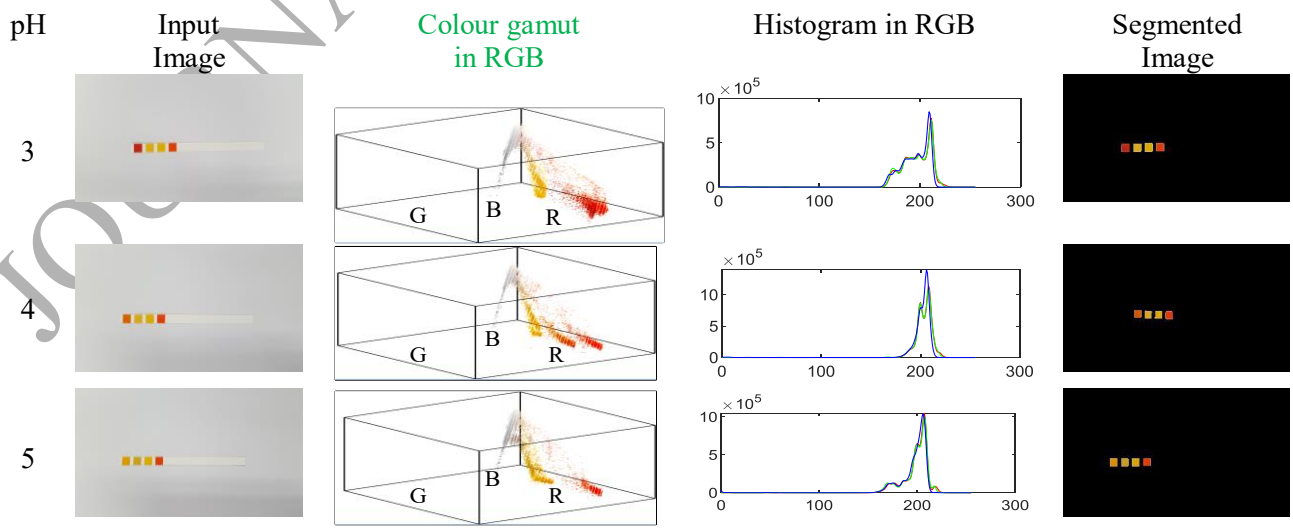
x_g = 8-bit unsigned integer array of filtered image in green channel

x_b = 8-bit unsigned integer array of filtered image in blue channel

The examples used in this paper, both universal pH strip and urine dipstick contain multi-objects per sample. In our original dataset, the relevant colours in the selected paper strip can be placed in 5 clusters or groups in theory.

$$g(x,y) = \begin{cases} g_1 \;\; if\; f(x,y) = foreground\; pixel \\ \qquad (colour\; pad\; 1) \\ g_2 \;\; if\; f(x,y) = foreground\; pixel \\ \qquad (colour\; pad\; 2) \\ g_3 \;\; if\; f(x,y) = foreground\; pixel \\ \qquad (colour\; pad\; 3) \\ g_4 \;\; if\; f(x,y) = foreground\; pixel \\ \qquad (colour\; pad\; 4) \\ g_0 \;\; if\; f(x,y) = background\; pixel \end{cases} \quad \text{..... (Eq. 1)}$$

In order to separate these clusters, clustering algorithm such as k-means, k-medoids and c-means can be used. However, as explained earlier (Fig. 3), for rapid execution, the relevant colours should be categorised as either foreground region of interest (ROI) or background colours (e.g. paper, tricyte).



JOURNAL PRE-PROOF

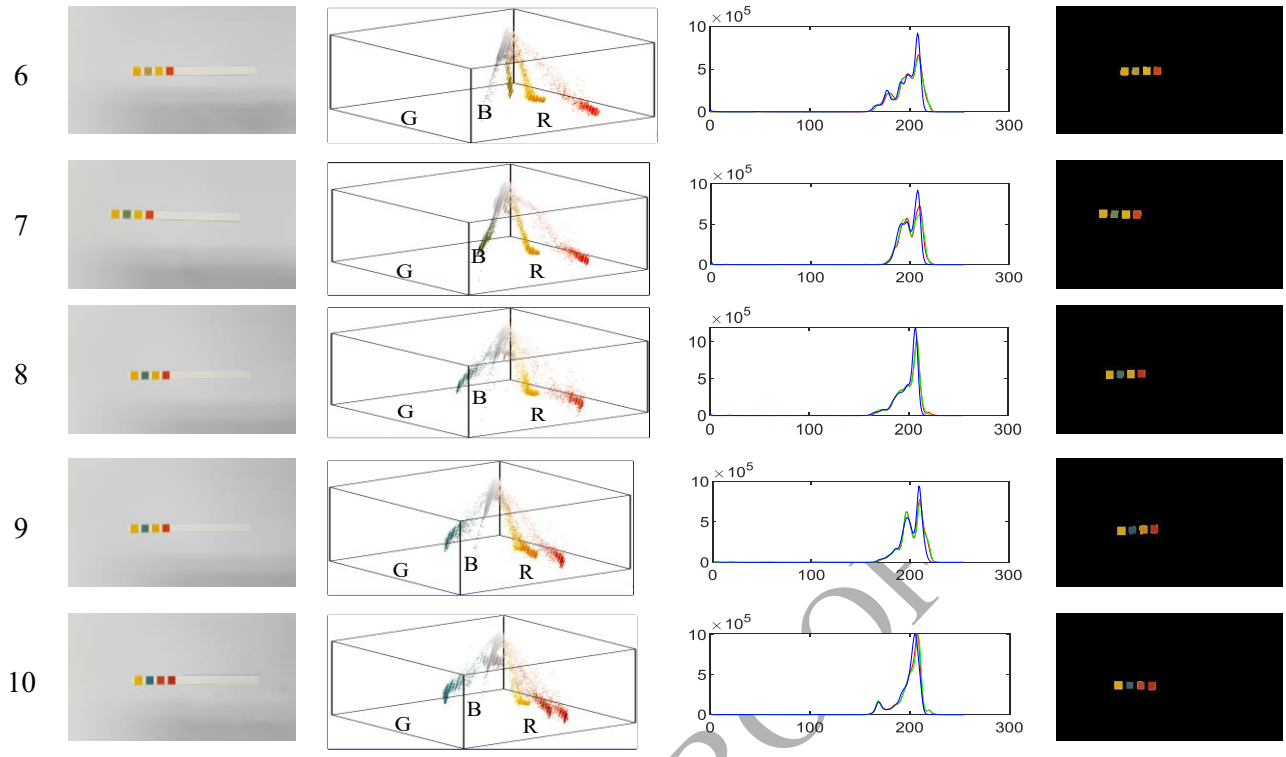
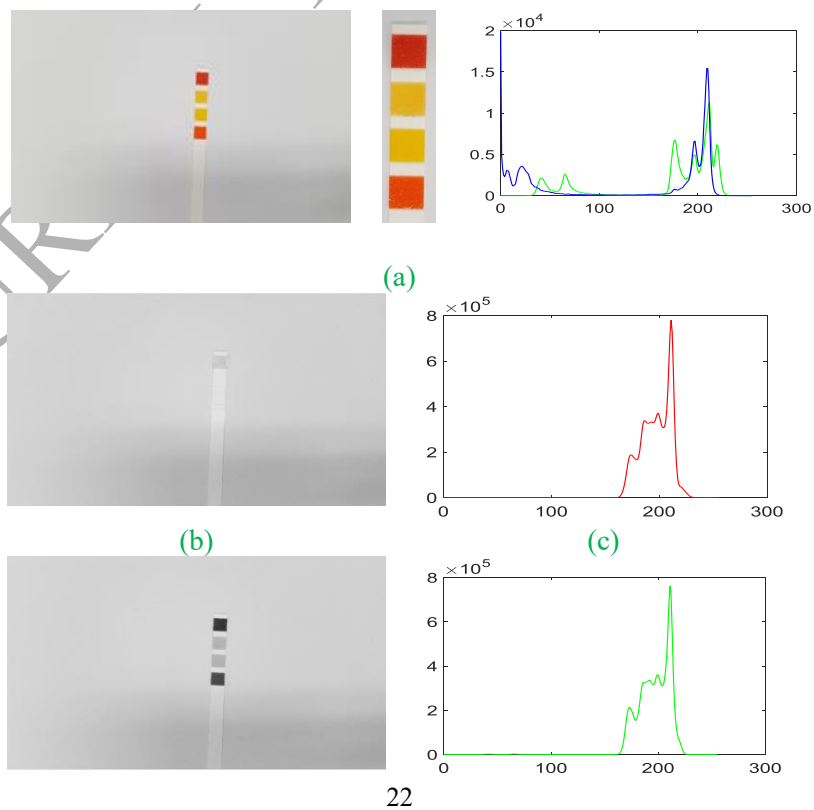


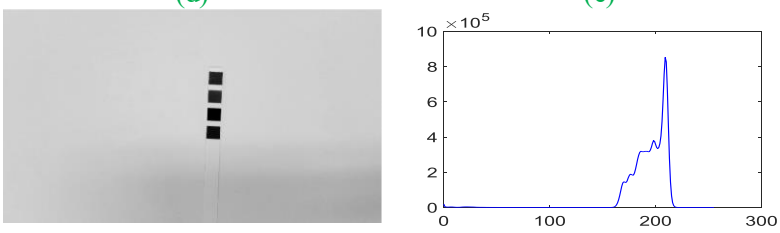
Fig. 7: Histogram based image segmentation

The colour gamut of the presented case for all class labels are shown in Fig. 7 in the form of point clouds for better visualisation of the colour distribution. It is difficult to perceive the content in the R, G and B channel from the colour histogram in Fig. 7. Therefore, an example of the contents of the background and foreground in all three channels is shown in Fig. 8.



JOURNAL PRE-PROOF

(d) (e)



(f) (g)

Fig. 8: (a) paper strip tested on pH 3, zoomed ROI and histogram of the zoomed image in green and blue channel, (b) input image 'a' in red (R) channel, (c) histogram of R channel, (d) image 'a' in green (G) channel, (e) histogram of G channel, (f) image 'a' in blue (B) channel, (g) histogram of B channel

For better visualisation, Fig. 8(a) illustrates one of the input images (class label: pH 3) along with its magnified ROI. Fig. 8(a) also contains the histogram of green and blue channels of the magnified ROI for the better conception of the intensity vs. number of pixels. When the channels are separated as it is mentioned in the framework (Fig. 5), the R channels holds both background as well foreground pixels, shown in Fig. 8 (b) and (c). The red components of the input image appear brighter in the red channel, as exemplified in Fig. 8(b), which is not sufficient to effectively distinguish the foreground colours from the background.

From the colour histogram analysis (Fig. 7 and Fig. 8), it can be observed that the higher band of G (green channel) and B (blue channel) holds the background pixels from g_0. Due to the type of the case study, the threshold level of the green channel overlaps with the foreground pixels of one of the four groups in Eq. 1. From Fig. 8(f), it is clearly visible that the background is better perceivable in blue channel, as the brightness of the foreground was significantly low. Therefore, the background separation can be performed more convincingly in blue channel.

The intensity of x_b was then normalised by Otsu (1979), prior to converting it into a binary image. Without the colour channel separation, the Otsu is not adequate enough to provide a reproducible performance. This issue can be visualised in Fig. 4. This paper explored several other alternatives, as shown in Fig. 3 and Fig. 4, which either resulted in a higher computational complexity or less reproducibility that are more sensitive to imaging condition. Therefore, this paper emphasises on the presented framework (Fig. 5). The step-by-step output of the framework is exemplified with an input image of class label pH 3 in the Supplementary Document 3.

JOURNAL PRE-PROOF

In continuation of the proposed framework, at this point after the binary conversion, the processed image only contained g_0. The logical negative of the matrix would provide the foreground pixels.

To extract all the blocks of the colour pad, Euler number property was utilised.

I= Binary foreground pixels

$O_i(n)$ = Objects in I by 8-connectivity, where $n = 1, 2, \ldots N$

h_o = Number of holes in $O_i(n)$

$e_{L,H}$ = $number\ of\ O_i - number\ of\ h_o$

C = Connected components by 2-D Euler Number ($e_{L,H}$)

All the pixels in C lower than the threshold value were eliminated to remove the noise in the binary image. Therefore, the image contained the pixel position of ROI only. As it is mentioned earlier, the main focus of this paper is the assays that contain multi-objects per sample where the number of objects per sample would vary from assay to assay. In our original dataset as well as the $א\_pH$ dataset, the universal pH test papers hold 4 colour pads and the decision of the pH level depends on the combination of these colour pads. Therefore, an object counting rule was deployed to increase the reliability of the system, discarding the false positives.

I_c= 8-bit unsigned integer array from a logical array containing C

$$Y = \begin{cases} x_r \times I_c \\ x_g \times I_c \\ x_b \times I_c \end{cases}$$

The final output obtained by the AND gate operation can be expressed as Y and illustrated in Fig. 7 as the segmented image.

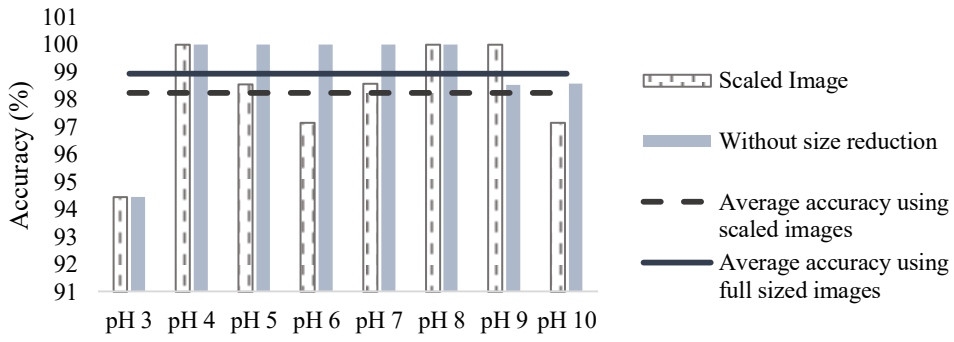


Fig. 9: Performance of the image processing algorithm

The performance of the algorithm or framework (Fig. 5) is illustrated in Fig. 9. The overall image processing performance showed >98% accuracy. The accuracy of the image processing algorithm is calculated based on Table S3. 1 (Supplementary Document 3). Considering the

fact that the light source '5' in Fig. S1.2 (Supplementary Document 1) varied during winter and spring in the UK while the images were taken, the image processing method exhibited good accuracy and reliable for the indoor environment.

After image segmentation, due to rule-based noise filtering, false positive ROI area was minimum. The image processing framework showed >97% accuracy for pH 4-10. Among the pH levels, the algorithm showed comparatively poor performance on pH 3. Further investigation revealed that the light source of Eq. S1.2 (Supplementary Document 1) had more variation on the days when dataset of pH 3 was generated, which created more shadows on the images. The position of the sample plate (Fig. S1.2, Supplementary Document 1) was kept constant for all the experiments. There were shadows due to the ambient objects that could have been avoided by moving the sample plane based on '5' (Fig. S1.2, Supplementary Document 1).

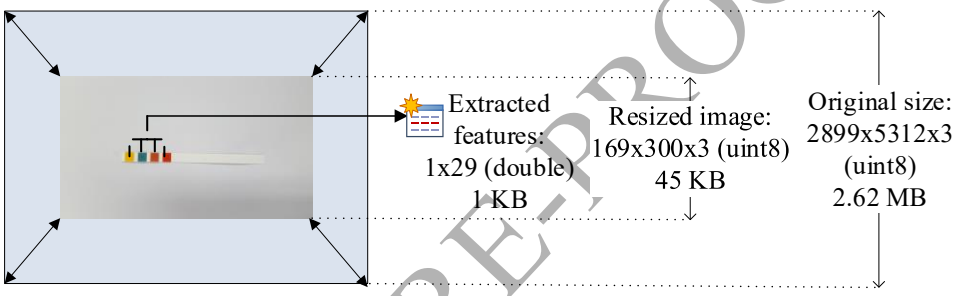


Fig. 10: Example of dimension reduction at different stage

In this work, we also evaluated the possibility of performance degradation of the histogram based image processing algorithm due to scaling operation tested on MATLAB in desktop environment. Without scaling and keeping the rest of the framework same, the full-sized images provided 98.94% accuracy on average. If the framework (Fig. 5) is followed, then the mean accuracy of the image processing algorithm was 98.23% (Fig. 9). Therefore, it can be stated that the dynamic scaling had a negligible impact on the image quality. There is an emphasis on scaling in the literature for mobile-enabled medical image processing (Bourouis et al., 2014). As the image processing algorithm in this work was not affected by resizing, after applying dynamic scaling, the maximum dimension of the image was 300 on average. The dimension reduction at a different stage, without compromising the performance, effectively reduced the overall memory occupancy of the system (Fig. 10), making it suitable for the mobile environment.

While the testing framework (Fig. 5) for the extended datasets, as shown in Fig. 11, the proposed hybrid image processing algorithms were found to be robust enough to successfully separate the ROI of א_pH papers without any additional fine-tuning.

JOURNAL PRE-PROOF

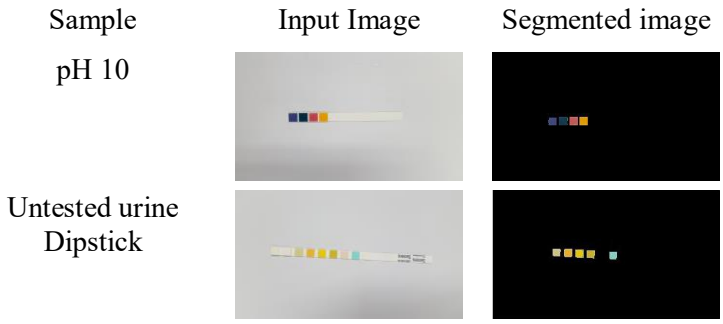


Fig. 11: Image processing of $\aleph\_pH$ paper using pH 10 buffer solution and untested urine dipstick

In order to further examine the adaptability of the image processing algorithm, another example (Multistix® GP, Siemens) of lateral flow assay was utilised, consisting 8 pads for different indicators (such as glucose, ketone, pH). There are two additional blocks for reference. These pads or blocks have similar length and width as our original dataset. However, the thickness and block-to-block distance are different. Moreover, the base is made of trycite, which are more hydrophobic than paper and has different reflectance. These test strips are multi-objects/ sample assays, however, each of these objects is designated or targeted for different test (e.g. pH, glucose) performed on the same sample (e.g. urine of a subject). Although each image contains multiple objects for a single test strip, they are single object/ target, known as multiplex assay. As illustrated in Fig. 11, the image processing framework (Fig. 5) was successful to separate the ROI of the multiplex assay of urine dipstick.

Different illumination conditions can influence the performance of an image processing algorithm (Smith et al., 2014). However, the proposed image processing framework showed consistent performance for D-A_i, D-jl and D-dl. The framework provided 353 correct segmentation out of 360 images with 98.06% accuracy.

JOURNAL PRE-PROOF

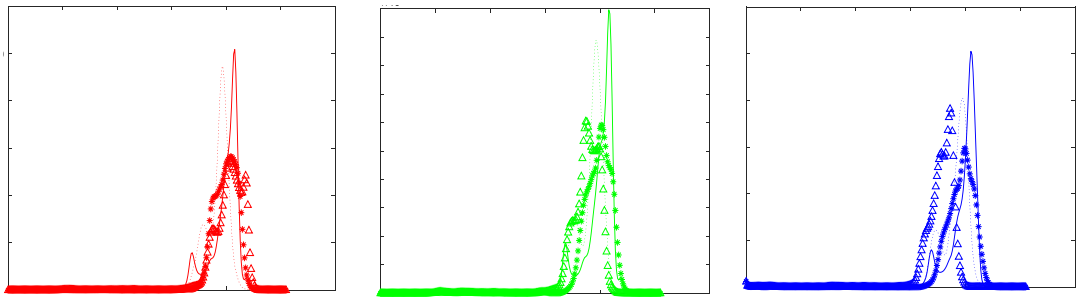


Fig. 12: Histogram of an image of a pH test strip under different illumination across three colour channels. In the colour histogram, the horizontal and vertical axis represents the intensity and the frequency or number of pixels, respectively.

The same pH test strip under different illumination conditions is visibly showing different histogram pattern using the same mobile phone camera in Fig. 12. The histograms in Fig. 12 confirm the need to include diverse illumination condition to develop a robust system, which is supported by the reported articles as well (Kim et al., 2017a; Solmaz et al., 2018).

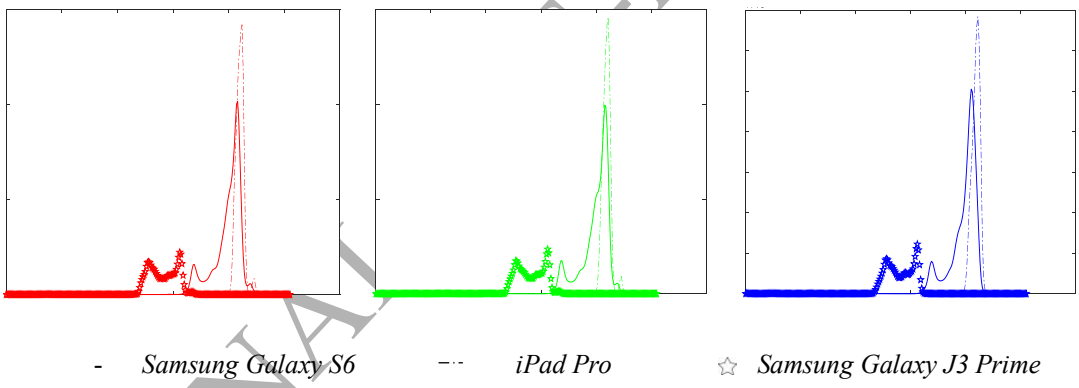


Fig. 13: Histogram of an image of a pH test strip across three colour channels using different mobile phone cameras. In the colour histogram, the horizontal and vertical axis represents the intensity and the frequency or number of pixels, respectively.

In Fig. 13, the impact of capturing an image of the same pH test strip using same illumination condition can be observed while incorporating the camera of a number of personal devices. Comparing the colour histogram of Fig. 12 and Fig. 13, it appeared that the images can be more affected by the illumination condition than the variation of the mobile devices.

In addition to Samsung Galaxy S6 (original dataset), the images were captured by two other devices: iPad Pro and Samsung Galaxy J3 Prime. The sizes of these devices are different, which effectively varied the exposure plane shown in Fig. S1.2 (Supplementary Document 1).

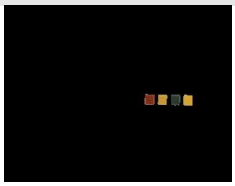
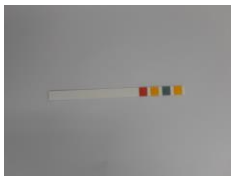
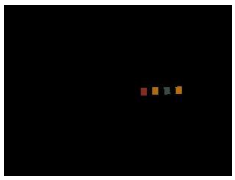
Device	Input Image	Segmented Image
iPad Pro		
Samsung Galaxy J3 Prime		

Fig. 14: Image segmentation by different devices

Therefore, the focal length of these devices had to be adjusted for the consistency of the overall system. The dimension of Samsung Galaxy J3 Prime is similar to Samsung Galaxy S6. Therefore, no adjustments were conducted on the focal length for this device. The iPad Pro was held at 11 inches away, varying the 35mm focal length within 66-68. It should also be mentioned that the images captured by the different devices varied in size due to the system-defined standardisation. Nevertheless, the image processing algorithm (Supplementary Document 2) showed consistent performance (Fig. 14) to segment the images captured by different devices. It was also observed that the presented system is capable of handling such variation in the focal length.

Due to the optimised intelligent components of the image processing algorithm (Supplementary Document 2, the image processing steps presented in this paper was initiated with the input image as it is captured, rather cropping the ROI to such as Mutlu et al. (2017) or time consuming object detection process such as Yetisen et al. (2014) as a separate step and consequently feeding the rest of the algorithm to analyse the profile that passed in the centre of the paper strip, or a smaller ROI. This alternative course of action than Mutlu et al. (2017) and Yetisen et al. (2014) aided in reducing the computational complexity without compromising the performance (Fig. 9). As mentioned earlier, 'performance' is defined by the evaluation criteria including execution time and operability within resource-limited settings. Integration of a step to consider a smaller ROI or only the centre of the strip would have obligated the system to the certain assays and reduce flexibility to be adapted to a different exposure of the ROI or variation in the shape, type, location and relative position of the objects within an assay. Evidence of the claim is demonstrated in Fig. 11, which could be further replicated, enhanced and validated for other paper-based assays.

JOURNAL PRE-PROOF

6. Classification and Regression of the Segmented Images

6.1 Feature Extraction

Once the ROIs were segmented, the characteristics of these samples were analysed from its colour moments. In our earlier study (Abuhassan et al., 2017), we have utilised lower order colour moments, c_m (Sergyan, 2008) in only LAB colour space (c_s) for all the colour channels (c_c) to classify wet-chemical based qualitative colourimetric tests. For stable paper assays such as pH indicator strips and urine dipsticks, the required feature-set may vary. Therefore, the impact of different attributes is required to be analysed. Let's assume, the required features for colourimetric tests can be expressed as Eq. 2.

$$ft = \{ft_{cs}(c_m) \times ft_{cs}(c_c)\} + ft(\Delta E_{LAB}) \quad \text{..... (Eq. 2)}$$

The feature-set in Abuhassan et al. (2017) can be described as Eq. 3, where 6 colour moments were considered in L, a and b channel, discarding the entropy in L channel. As the images involved single objects per sample, there were 17 features in total.

$$ft = ft_{LAB}(c_m) \times ft_{LAB}(c_c) - Entropy_L \quad \text{..... (Eq. 3)}$$

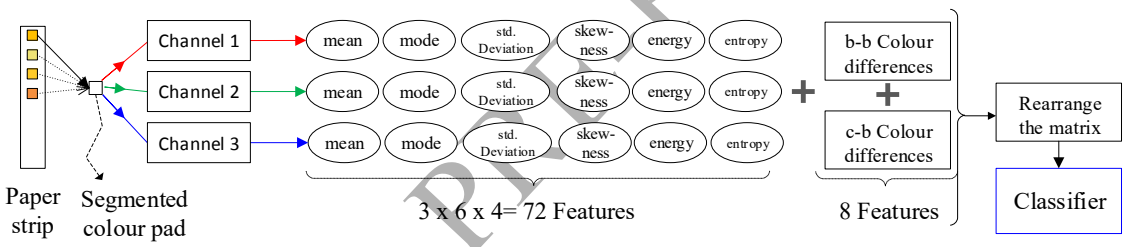


Fig. 15: Feature extraction framework

The feature extraction framework is illustrated in Fig. 15. In this work, the colour differences in LAB colour space are calculated in a closed loop as additional features (Tania et al., 2017). If the colour block is b_p and position of the block is p; $p \in Z^+$, then let's calculate the colour difference for each block calculated from its previous block using the following equation.

$$\Delta E^*_{b_p} = \sqrt{\{(mean\ L^*_{b_p} - mean\ L^*_{b_{p-1}})^2 + (mean\ a^*_{b_p} - mean\ a^*_{b_{p-1}})^2 + (mean\ b^*_{b_p} - mean\ b^*_{b_{p-1}})^2\}} \quad \text{..... (Eq. 4)}$$

The features generated from Eq. 4 are novel features, where a pseudo-control colour set is generated for each individual test using the user input image itself. However, in real life situation, one would require to compare the tested paper assay with a colour chart. This colour chart holds the control colour for the block or sample. In the literature, the colour of the block is often tracked against the colour chart (Rahmat et al., 2018; Solmaz et al., 2018). In this

JOURNAL PRE-PROOF

work, we have also used the colour difference of each colour block from the corresponding control colour block (c_p) as part of the feature-set.

$$\Delta E^*_{c_p} = \sqrt{\{(mean\ L^*_{c_p} - mean\ L^*_{b_p})^2 + (mean\ a^*_{c_p} - mean\ a^*_{b_p})^2 + (mean\ b^*_{c_p} - mean\ b^*_{b_p})^2\}} \quad \ldots\ldots\ldots \text{(Eq. 5)}$$

In Eq. 2, $ft(\Delta E_{LAB})$ was obtained from Eq. 4 and Eq. 5, which could be stated as Eq. 6.

$$ft(\Delta E_{LAB}) = \Delta E^*_{b_p} + \Delta E^*_{c_p} \quad \ldots.. \text{(Eq. 6)}$$

The case study involves multi-objects for single sample per image, which elongated the feature set. Therefore, initially, 440 features[5] in total were considered to train the classification model.

6.2 Classification and Feature Analysis

6.2.1 Feature Selection

Understandably, the most important features for a colourimetric test would be colours, which was reflected in the reported articles as well (Table 4). An inadequate feature-set can lead towards under-fitting, whereas an elongated one would result in a higher dimension of the feature-set which would cost the system in higher computation time and occupying more memory space of the personal devices.

JOURNAL PRE-PROOF

[5] Eq. 4: No of features: 1 x4=4; Eq. 5: No of features: 1 x4=4; Eq. 2: (6 colour moments x 4 blocks x 3 colour channels x 6 colour spaces)+ Eq. 6= = 432 features + Eq. 6; Eq. 6: No of features 4+4=8. Therefore, Eq. 2= 432+8= 440 features

Table 4: Feature analysis methods

Attribute	Description	Reference
Colour moment	Mean	Rahmat et al. (2018); Mutlu et al. (2017)
	Multiple	Kim et al. (2017a)
Colour difference	Initial ~ end point	Wang et al. (2016); Vashist et al. (2015)
Colour space	RGB	Rahmat et al. (2018); Mutlu et al. (2017); Soni and Jha (2017); Wang et al. (2016b); Sicard et al. (2015)
	HSL	Akraa et al. (2018)
	HSV	Lopez-Ruiz et al. (2014)
	LAB	Alankus et al. (2018); Konnaiyan et al. (2017)
	Multiple	Kim et al. (2017a)
Grey value	Weighted mean	Khan and Garnier (2013)
	Green channel	Barbosa et al. (2015)
	Rule-based	Sicard et al. (2015)

The conventional semi-quantitative test via RDT often offers a colour chart, for example, the colour chart for urine dipstick. Rahmat et al. (2018) provided the colourimetric measurement using only ΔE calculation. On the other hand, Vashist et al. (2015) provided the colourimetric decision by plotting the pixel intensity against the concentration of the analyte. Konnaiyan et al. (2017) utilised correlation graph. The use of PCA is also quite prevailing (Akraa et al., 2018). Jonas et al. (2016) performed a colour comparison from its ratio. Garg et al. (2014) utilised a calibration curve to determine the concentration of glucose. Although a colour chart was available for the dipstick used by Shen, Hagen and Papautsky (2012), the study used a calibration curve to measure the pH level. A binary decision for substance tracing in Smith et al. (2014) was availed from the known cocaine standards.

On the other hand, this work performed feature expansion to consider a wide range of potentially important and relevant features, mostly signal features, before performing any

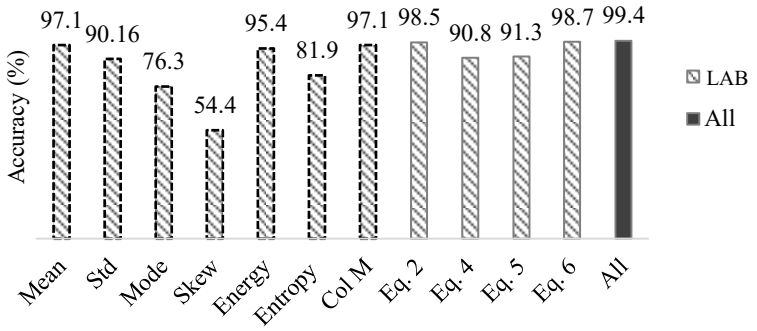
JOURNAL PRE-PROOF

feature selection and optimisation. At first, the classifiers were trained using 440 colour features. Among the standard classifiers including discriminant analysis, support vector machines (SVMs), k-nearest neighbours (KNNs) and ensemble methods, the ensemble method called Subspace discriminant analysis showed the best performance. Using 30 base learners and subspace dimension of 220, the classifier provided 99.4% accuracy with three misclassifications. The training was conducted in 33.418 s and the prediction speed was 350 observations/s. However, more features would mean more model complexity and requirement of more storage and processing capacity. Therefore, the number of features is required to be optimised. The optimisation was conducted using univariate analysis at different stages. For the convenience of faster training, the analysis was executed via MATLAB classification learner application.

The use of histogram features in the reported articles can be observed in Table 1. This paper systematically investigated the impact of the histogram features including colour moments and colour spaces. Additionally, the impact of control colours was studied as well. We have explored the colour moments in LAB colour space to begin with.

Based on the univariate analysis conducted on $ft_{LAB}(c_m)$, the best performing features were found to be mean and energy (Fig. 16). These two features comprise good signal, performing as proficient (97.1% accuracy) as the combined features of $ft_{LAB}(c_m)$. Considering the fact that we are analysing the colourimetric tests, the average colour or brightness of the ROI is a key feature. The mean colour value is considered to be the most important features in the reported articles as well (Mutlu et al., 2017; Rahmat et al., 2018). On the other hand, the energy in L, a and b are the amplified brightness level. Therefore, energy is directly linked with the performance of the mean colours. Due to the reflectance on the two dimensional surface, the unbiased standard deviation within each colour pad did not provide significant contrast to improve the classification performance. The mode of colours on the solid surface of individual ROI without any opto-mechanical attachment can be misleading. Because, the system would be susceptible to the ambient lighting environment. The mode is supposed to provide effective information, if the ROI is scaled such as hardware systems of CLINITEK Status® + Analyzer and the wavelength is filtered (Siemens Healthcare GmbH, 2018). On ideal condition, there should not be any asymmetry of colour distribution, therefore the SKEW can also be discarded from the feature-set. Hence, among the features from $ft(c_m)$, only mean and energy were considered to train the classifier.

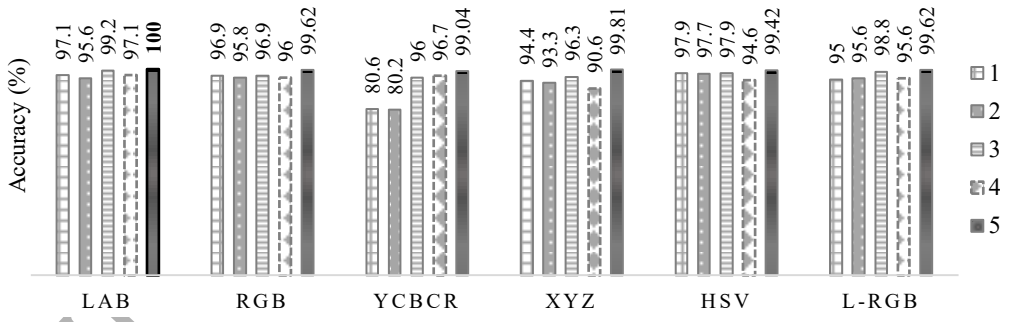
JOURNAL PRE-PROOF



Std: Standard deviation; Col M: Colour Moments; All: 440 features in 6 colour spaces. The features with dotted outlines are functions of $ft_{LAB}(c_m)$

Fig. 16: Performance of the features

As illustrated in Fig. 16, the control colours were also found to be influential features. Only the control colours are sufficient enough to provide a 98.7% accurate classification using Eq. 6, which was studied further while exploring and fine-tuning the classifier. Similar to our earlier work (Tania et al., 2017) on the wet-chemical based colourimetric test, the colour differences (Eq. 4-6) are considered in LAB space only. The L, a and b, imitating the nonlinear response of the human eyes, can also resemble the uniform changes in perceived colour facilitated by the uniform changes in the LAB- components. Therefore, the control colour related calculations were computed in a colour space which has more advantages at Euclidean space.



1: Mean; 2: Energy; 3: Mean, energy and Eq. 4; 4: Colour moments. 1-4 are the performances by the best performing classifier excluding LS-SVM; 5: LS-SVM using Mean, energy and Eq. 4

Fig. 17: Performance of colour spaces

After the initial assessment of the colour moments and control colours as features, the colour spaces were appraised (Fig. 17). H. Kim et al. (2017) utilised four colour spaces to provide a colourimetric decision, whereas Solmaz et al. (2018) used three (Table 1). Mutlu et al. (2017) used RGB images in JPEG along with capturing the images at the RAW format. The standard devices capture the image in sRGB colour space. In this paper, we have studied the performance of six colour spaces on the original dataset. Both HSV and LAB are closer to

human colour perception. HSV separates the intensity from the colour information. Therefore, for a robust system, HSV can help to deal with the lighting condition and shadows. Similarly, the 'a' and 'b' channel of LAB colour space signifies the colour. HSV is a cylindrical model, which gives the LAB more advantage over HSV. The linearisation of linearised gamma-corrected RGB (L-RGB) was conducted using the sRGB standard. Among the colour spaces, the LAB was found to be the most influential colour space. The strength of the LAB is the perceptual uniformity property.

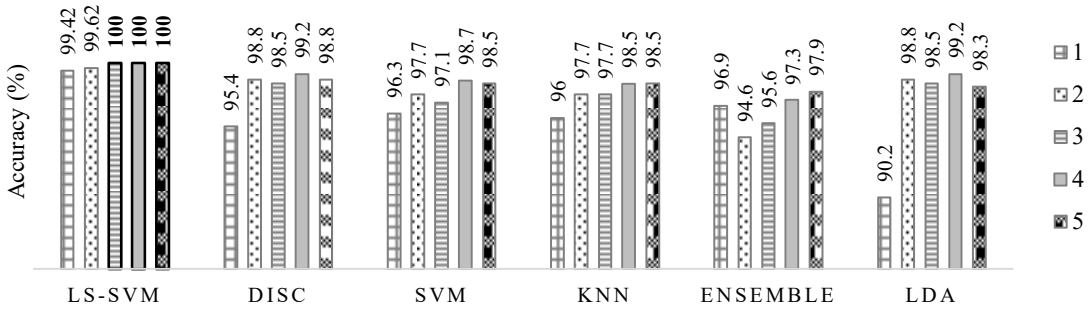
Another popular dimension reduction technique, principal component analysis (PCA) combining correlated attributes to create superior new features did not improve the overall performance. Therefore, based on the performance of different features and colour spaces, mean, energy and the control colours in three channels of LAB were chosen to be feature-set (32 features) to explore the performance of the classifiers.

6.2.2 Performance of Classification Methods

Exploring 440 features, the selected 32 features were identified as the good features to ensure that the classifiers are trained with signals, not noise. Different supervised learning techniques were evaluated to provide the semi-quantitative colourimetric classification. The list of classifiers includes LS-SVM, LDA and RF which provided good accuracy in the reported articles for similar classification task (Mutlu et al., 2017; Solmaz et al., 2018; Dhar, Mehta and Sit, 2017). The classifiers were trained and 10-fold cross-validated by the original dataset, followed by a re-evaluation of the good features.

After careful selection of 32 features, SVM, KNN and discriminant analysis exhibited similar performance in Fig. 18 (>98% accuracy). The overall performance from the discriminant analysis was good. For a different combination of the good features, the LDA and quadratic discriminant analysis (QDA) outperformed each other. Therefore, subspace discriminant also exhibited good performance as an ensemble method. The performance of Random Forest was also notable. However, the LS-SVM (Suykens and Vandewalle, 1999; Suykens et al., 2002) showed the best performance using the selected performance as illustrated in Fig. 18. Due to the weighted function with a modified cost function, it is more robust than SVM. The performance of LS-SVM was consistent for any good features in any colour space (Fig. 17 and Fig. 18).

JOURNAL PRE-PROOF



1: Mean; 2: Mean and energy; 3: Mean, energy and Eq. 4; 4: Mean, energy and Eq. 5; Mean, energy and Eq. 6

Fig. 18: Performance of different classifiers

After identifying the best performing classifier using good features or signals, the selected 32 features were explored again which revealed the features can be further reduced to 28. The selected 32 features, including both feature-set of control colours (i.e. Eq. 6), LS-SVM provided 100% accuracy. As shown in Fig. 18, in addition to mean and energy, the use of only one feature-set of control colours i.e. either Eq. 4 or Eq. 5, is capable to provide 100% accurate colourimetric classification for LFA. Therefore, instead of Eq. 6, only one set of control colours can be used.

In order to specify between the choices of control colours, this paper suggests, it is better to use the pseudo control colours (Eq. 4), especially for multi-object single-target colourimetric tests. The control colours (Eq. 5) generated from the colour chart are the features generated one time, acting as a ground truth may widely vary from the condition where the user input image is captured. Thus, the user input can appear as a noisy image. The pseudo control colours (Eq. 4) are generated each time using the user input itself. Therefore, these features are more reliable.

From

JOURNAL PRE-PROOF

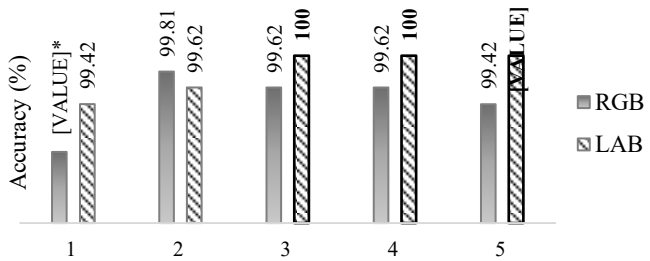
Table 5, it can be observed that the specificity and sensitivity achieved by LS-SVM is also 100%, which are our key evaluation criteria. Among the reported articles in Table 1, Mutlu et al. (2017) performed the colourimetric classification utilising the same application i.e. paper-based pH test strip. Similar to Mutlu et al. (2017), this paper also found LS-SVM to be the best performing classifier for the presented classification problem. Mutlu et al. (2017) showed that, with and without apparatus, the system can exhibit the same performance. The experiment of Mutlu et al. (2017) without any apparatus included 270 images to provide the classification of pH strips, which explains the confidence interval (CI) percentage in Table 5.

JOURNAL PRE-PROOF

Table 5: Performance of LS-SVM

Ref.	Number of classes	Number of samples/ class	Sensitivity/ class (%)		Specificity/ class (%)		Accuracy (%)	
			Value	95% CI	Value	95% CI	Value	95% CI
Mutlu et al. (2017)	15	18	100	81.47-100	100	98.56-100	100	98.66-100
This work	8	65	100	94.48-100	100	99.19-100	100	99.29-100

Mutlu et al. (2017) utilised images saved in different file formats which increased the volume of the data, however, did not carry any significance in terms of features or classifiers. Therefore, the effective feature-set of the experiment of Mutlu et al. (2017) can be considered as the mean colours at RGB colour space. Using the feature-set of Mutlu et al. (2017) on our original dataset, the performance degraded from 100% to 98.85% accuracy (Fig. 19), which can be perceived from the data generation point of view. Due to the light source '5' (Fig. S1.2, Supplementary Document 1) and use of 65 independent test strips for each class, the original dataset contains much robust data, whereas Mutlu et al. (2017) used the same pH test strip at different orientation using 3 different light sources. The same 12 features, i.e. mean colours in LAB colour space showed better performance due to its strength of colour separation and handling shadows. The RGB is good at modelling the output of the phone camera, but LAB is closer to the 'human colour perception' and the presented problem deals with mimicking the naked-eye measurement of the colourimetric tests using computer vision. Therefore, the LAB is certainly a better choice of colour space.



Using original dataset- 1: Mean ; 2: Mean and energy; 3: Mean, energy and Eq. 4;
4: Using new_pH10 and 28 features (mean, energy and Eq. 4);
5: Using $א\_pH$ and 28 features (mean, energy and Eq. 4)
*Using the feature-set of Mutlu et al. (2017)

Fig. 19: Comparative performance of selected featured in LAB and RGB using LS-SVM

To further explore the performance of the classifier as well as the optimised 28 features on unseen images and validate the system using a similar lateral flow assay, we extended our experiment using new_pH10 and $א\_pH$ from Table 2. The extended experiment can validate

the reliability, adaptability and robustness of the system. The new_pH10 contains 65 images of pH test strips of level 10 at various random orientation. The classifier was not effected by more variation in the orientation, as it provided 100% accuracy using 10 fold cross validation (Fig. 19).

In order to validate the system's adaptability to similar lateral flow assays, $א\_pH$ dataset was utilised. A similar dataset is available in GitHub (Dhar, Mehta and Sit, 2017). However, there are only 10 samples per class available in the open source domain. The pH indicating colours of $א\_pH$ dataset are different than our original dataset, therefore the classifier was re-trained using the optimised 28 features, keeping rest of the hyper-parameters same. The 10 fold cross validation showed consistent result with 100% accuracy (Fig. 19).

Using a pH indicator paper of different brand ($א\_pH$ dataset) effectively changed the colours of each block for the same class label. There was a slight variation in the block size and block to block distance as well. As the performance of the system was as good as the original dataset, it justifies our choice of the classifier as well as the feature set, confirming the reliability and adaptability of the system.

After finalising the classifier, optimising the features and cross-sectional performance evaluation, the extended dataset (D-lights) was utilised. As mentioned earlier, due to a significant shift in the colour histogram (Fig. 12 and Fig. 13), consideration of such diversity would enhance the reliability of the system. After including these 360 images with the original dataset, the optimised features and LS-SVM showed consistent performance for the dataset (D-lights). The effect was further analysed separately without including the original dataset (Fig. 19).

Using the optimised 28 features on all three datasets of paper-based universal pH test strips, the performance of the top performing classifiers were evaluated using statistical approach. The analysis is provided in the Supplementary Document 4, based on which it can be stated that LS-SVM outperformed the rest of the algorithms for all the dataset, which justifies our choice of the algorithm.

6.3 Regression

The qualitative colourimetric test can be seen as a pure classification problem, whereas the quantitative colourimetric test can be presented as a regression problem. The semi-quantitative colourimetric tests such as pH test can be described with a classification as well as a regression model.

JOURNAL PRE-PROOF

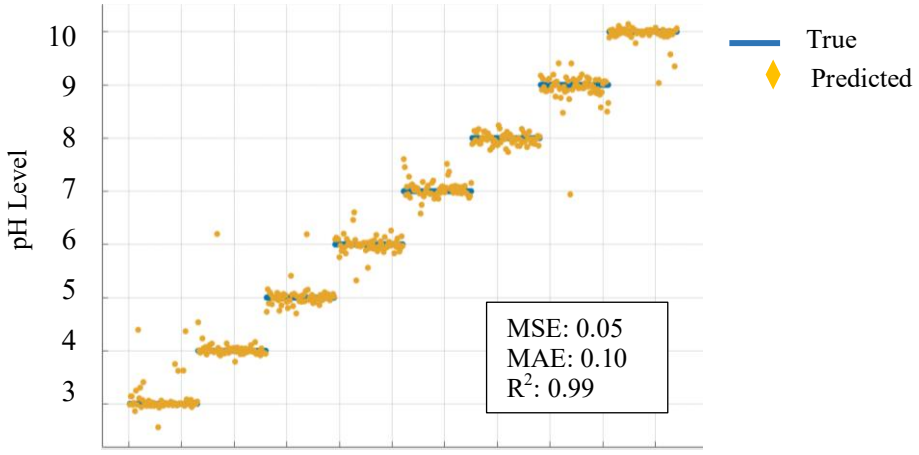


Fig. 20: Performance of Gaussian Process Regression. RMSE: 0.22

Among different regression models such as linear, support vector machine and ensemble methods, the Gaussian Process regression (GPR) (Rasmussen et al., 2006) showed better performance using the optimised 28 features (Fig. 20). The performance was compared by the square root of the mean squared error (RMSE). It is an estimation of the standard deviation of the error distribution. The coefficient of determination (R^2) value suggested that the model can explain approximately 99% of the variability in the selected response variables. It took 11.51 seconds to train the model and the prediction speed was approximately 7300 observations per second. We have utilised isotropic kernel. The covariance function, $k(x_i, x_j)$ estimates the course of response at point x_i effecting the response at a further point x_j, where x_i = predictor values, $i = 1 + 2 + 3 + \cdots + n$ and $i \neq j$. The Euclidean distance between x_i and x_j, $r = \{(x_i + x_j)^T(x_i - x_j)\}^{1/2}$. If y_j = target, σ_f = signal standard deviation, $\sigma_f > 0$ and σ_l = characteristic length scale, $\sigma_l > 0$, then the kernel function having same length scale for each predictor utilised in this work can be expressed as the following:

Covariance function, $k(x_i, x_j) = \sigma_f^2(1 + \frac{\sqrt{5r}}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}e^{\left(-\frac{\sqrt{5r}}{\sigma_l}\right)}$

In this work, as the system was trained on <1000 samples using the best performing features only, the GPR performed better than the other models due to well-balanced bias and variance, smoothing, optimised hyper-parameters and local generalisation. The kernel scale parameter attained the random basis for random feature expansion utilising sub-sampling based heuristics. The model may require further rectification using continuous pH levels (quantitative) as well as a larger dataset to substantiate the performance.

7. Deep Learning using Pre-trained Model

The deep learning has brought some recent success for image classification including object identification reducing the exasperation of the image processing (Krizhevsky, Sutskever and

JOURNAL PRE-PROOF

Hinton, 2012; Szegedy et al., 2015; He et al., 2015). However, in order to train the system using deep learning, it is essential to have a large dataset, and a powerful and high performing computational system. Therefore, in this work we have utilised transfer learning approach using pre-trained models such as AlexNet (Krizhevsky, Sutskever and Hinton, 2012), two versions of Inception (Szegedy et al., 2015) and ResNets (He et al., 2015) allowing us to train the model with less number of samples and faster computation. To the best of authors' knowledge, this is the first attempt to perform the colourimetric test using deep learning.

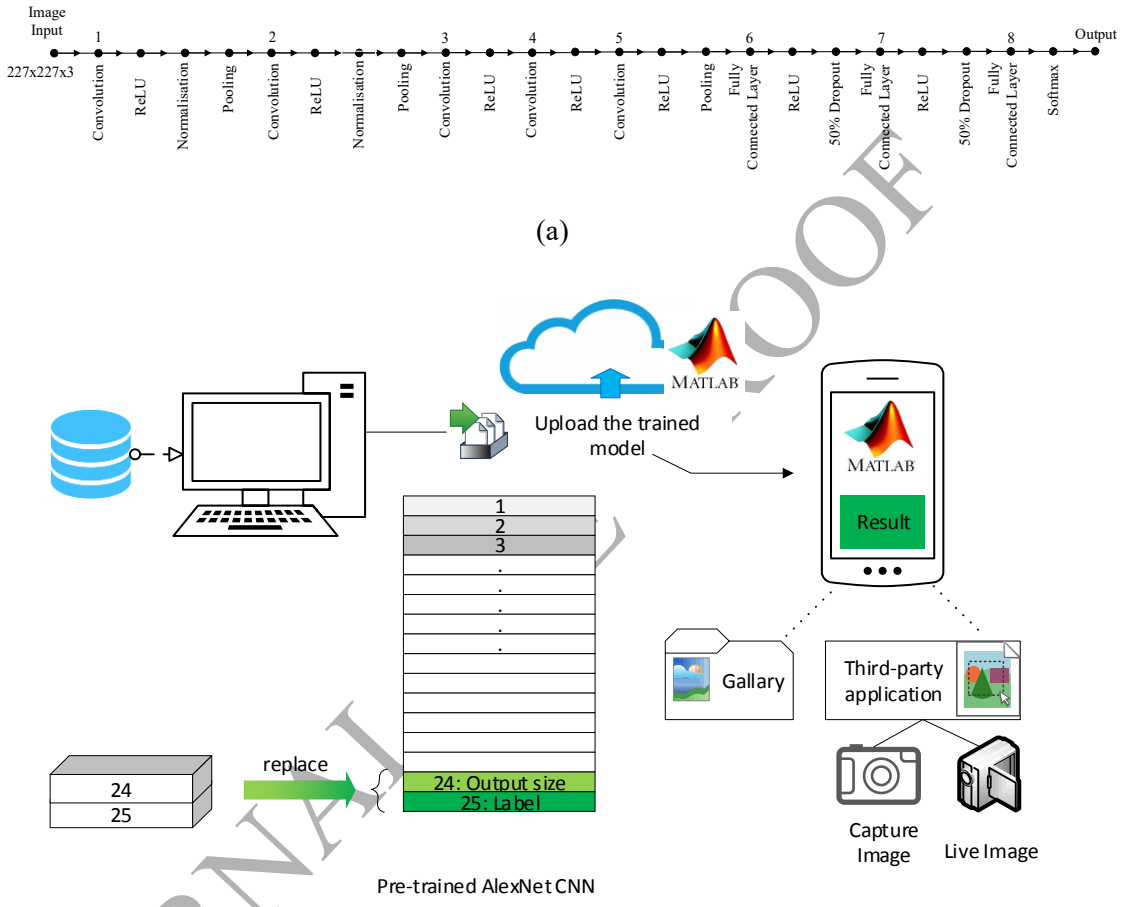


Fig. 21: (a) Layers of AlexNet and (b) Mobile-enabled server based pH test using AlexNet

At first, the input images require re-sizing to comply with the pre-trained models. After loading the pre-trained CNN, the final layers are replaced to fine-tune the model. An example is shown in Fig. 21. After training the model, the result can be pass on to the subsequent stages. In case of inception modules based GoogLeNet, the final three layers of the network have to be replaced to retrain the model with our dataset. These three layers are: 1000 fully connected layer called 'loss3-classifier', softmax and the output layer. These layers embrace the mechanism to train the model from the extracted features into class probabilities and

labels. In case of an advanced version of GoogLeNet i.e. Inception-v3, the final three layers are replaced and connected to the 313th connection that performs average pooling.

Similarly to GoogLeNet, the final layers has to be replaced to fine-tune the model of 50 layers deep ResNet-50 and 101 layers deep ResNet-101. However, unlike AlexNet, the learning rate of the intermediate layers of Inception models and ResNets were set to zero to minimise the training time and prevent overfitting. For example, the GoogLeNet comprises 22 layers with 144 connections. Computation of selected gradients is prevented by freezing initial 110 connections up to inception_5a module, effectively making the process faster.

Table 6: Deep learning based pH test using transfer learning

Model	AlexNet	GoogleNet	Inception 3	ResNet 50	ResNet 101
Elapsed time (min)	6.00	13.19	47.27	35.39	67.13
Size (MB)	668	67	232	259	456
Accuracy (%)	86.86	50	55.26	63.82	71.05

As it can be observed from Table 6, the training time was comparatively faster due to the smaller dataset as well as suspension of the learning rates of all the parameters of the earlier layers (apart from AlexNet). The memory size mentioned in Table 6, contains all the variables including the training and testing dataset. Considering the original dataset, 70% of the dataset was used for training and the rest for testing. The models were trained by a single CPU and then uploaded to the MATLAB server. An example is illustrated in Fig. 21, where the server based system was deployed on MATLAB Mobile. MATLAB mobile exploiting a third party application[6] to capture the new images on site.

The pre-trained model utilised in this work is relying more on the pseudo colours or the illumination rather the colour feature itself. Therefore, in the absence of distinct geometric features, the performance of Table 6 is perceivable. These convolutional networks require more fine-tuning, longer training cycle and larger dataset to further investigate regarding the poor performance by the pre-trained models.

[6] At this moment, MATLAB mobile cannot pull input image without using any third party application.

JOURNAL PRE-PROOF

Table 7: Training Time and Model Size

Algorithm	AlexNet	LS-SVM				
Features	--	12	24	28	32	440
Elapsed time (sec)	360	0.1087	0.1101	0.1162	0.1234	0.1941
Size (KB)	$198.8x10^3$	80	127	142	158	$1.228\ x10^3$

On the other hand, the best performing classifier using traditional machine learning techniques in this paper for the presented problem was found to be LS-SVM that does not require GPU, smaller in size to be deployed on mobile devices as a native application and can be trained using a smaller dataset with reliability and adaptability. One of the advantages of deep learning for image classification is subsiding the feature extraction and analysis steps at the data pre-processing stage. In order to generate a high accuracy using LS-SVM, the features required to be carefully selected.

Considering the existing resources and the demand for POC solutions, the use of deep learning does not comply with our evaluation criteria, i.e. ASSURED criteria. The trained model uploaded on the server, is bigger in size (Table 6) making it less suitable for the mobile devices. Using traditional machine learning techniques, Kim et al. (2017) showed that the stand-alone mobile application is two times faster than the server-based application for paper-based assays, which would worsen for larger models such as Table 6. The cloud operated system would require a more secure system in case of sensitive data such as health information.

8. Discussion

The presented intelligent system is consisting of both qualitative (visual primitives) and quantitative (signal features) nature of the visual knowledge. Exploring both traditional machine learning as well as deep learning approach, this paper suggests that traditional machine learning techniques both classification and regression are efficient enough to perform rapid, specific and sensitive colourimetric tests due to reliance on the colour histograms features only. The use of deep learning for colourimetric detection would be similar *'to break a butterfly upon a wheel'*.The colourimetric classification would require the pre-trained model to search for only the colours even when the rest of the geometric features are similar, and the occurrence of those colours are in the same location. Therefore, it is more logical to use a simpler machine learning model for the colourimetric classification, instead of

JOURNAL PRE-PROOF

building more deep layers which would require more processing capacity, memory size, larger dataset and more dependency on the cloud-based approach.

This work evaluates the performance of analogues computational systems for lateral flow assays using ASSURED criteria. In the absence of the exact dataset, in this work we have compared the performance based on the concept of recently reported high accuracy systems (Table 8).

Table 8: Comparative performance using ASSURED criteria

Reference	Application	A	S	S	U	R	E	D
H. Kim et al. (2017)	Alcohol saliva test	↓	PPV-NPV: >95%		↓	Robust, <30s	No	Yes
Solmaz et al. (2018)	H_2O_2	↑	N/A	N/A	↓	Robust, N/A	Yes	Yes
Mutlu et al. (2017)	pH test	↑	100%	100%	↓	Robust, N/A	Yes	Yes
Rahmat et al. (2018)	Urine dipstick	↑	>98.25%	>98.25%	↓	Not robust, N/A	Yes	Yes
This work	pH test	↑	100%	100%	↑	Robust, Real time	Yes	Yes

N/A: Information not available

As discussed in Section 2, the affordable systems are more likely to be equipment free and more accessible. In Table 8, the expense of the systems are comparative, e.g. although the additional hardware attachment is low-cost and straightforward (Kim et al., 2017a), it would be still more costly than a system which does not require such attachments at all. The presented system in this paper utilises the built-in camera of the smart devices such as mobile phone without enhancing or channelling the light with any additional hardware such as Kim, Awofeso, Choi, Jung, & Bae (2017), making the system convincingly more portable and easily operable.

The accuracy of the reported articles is presented in Table 1. The specificity and sensitivity (or similar metric such as precision-recall, type I- type II error, PPV-NPV) are often not

described in the articles. Based on the available information, a comparison is provided in Table 8. It should be also taken into account that the dataset of Table 8 varied in terms of a number of test strips, images of the same sample for repeatability and variation within the dataset. The data were often pre-processed manually or with an aid of additional hardware.

Using the original dataset, we have trained the better performing classifiers affirmed in the reported articles using the same feature-set as mentioned in Table 1. In the case of a few missing hyper-parameter values, we have utilised the default values in MATLAB. In Table 9, Rahmat et al. (2018) is excluded as it involves the same feature-set as Mutlu et al. (2017). Analysing the performance from these tables, the justification behind the choice of our classifier, i.e. LS-SVM is well supported by Table 1 and clearly evident from Table 9. Among the classifiers, artificial neural network (ANN) with 10 hidden layers (Kim et al., 2017a) showed a poor performance, which requires further investigation.

Table 9: Comparison of Accuracy (%)

Features	LDA	SVM	ANN	RF	Sub disc	LS-SVM
H. Kim et al. (2017)	Failed	97.7	68.8	97.9	99.6	100
	= 97.7					
Solmaz et al. (2018)	98.7	97.5	57.8	**96.3**	98.3	**99.62**
Mutlu et al. (2017)	88.1	96.9	64.2	91.2	76	**98.85**
This work	98.5	97.1	83.1	96.5	92.7	**100**

Both paper test strips of H. Kim et al. (2017) and Solmaz et al. (2018) has only one colour pad. Therefore, average binning of four colour pads could not aid the classifier in Table 9 while reproducing Kim's work using our dataset. H. Kim et al. (2017) utilised hardware attachment as well, which helped to discard noise from the colour signals. In this paper, using the feature-set and classifier on our original dataset, the attained accuracy was 97.7%, whereas the same feature-set provided a higher accuracy using LS-SVM as well as the ensemble classifiers (Table 9). The performance of the combination of the features of Solmaz et al. (2018) in Table 9 can be perceived from the performance of mean colours in different colour spaces in Fig. 16. A detailed comparison with Mutlu et al. (2017) is already presented earlier in Section 6.2.

In Table 8, different work included a different element of robustness. As described earlier, the robustness of the system can be represented by adaptability. In this work, we utilised

analogous separate data set to evaluate the adaptability of the image processing algorithm, selected optimised feature-set and classifier. The system was found to be adaptable to the new sample sets.

The elapsed time for training and prediction for the systems described the reported articles on our original dataset is shown in Fig. 22. A ranking in Fig. 22 is provided based on the computational cost and model size. A higher number of features would undoubtedly increase the size of the model. The size can also get affected by the complexity of the classifier itself.

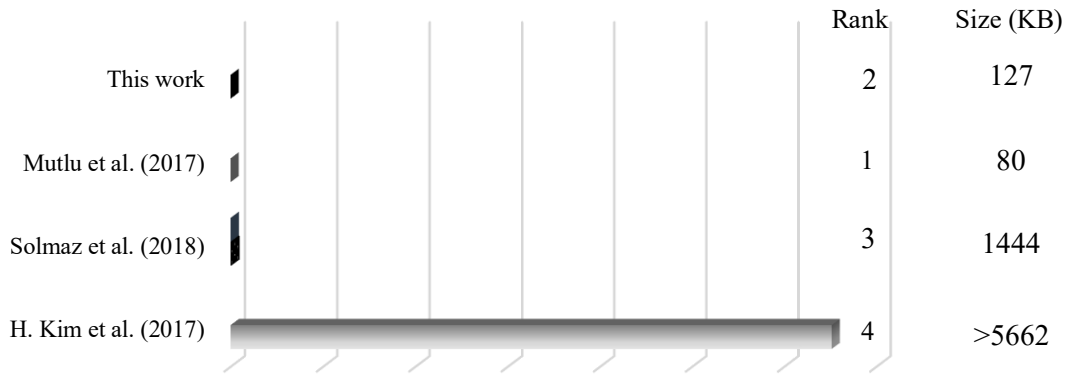


Fig. 22: Elapsed time in seconds. The rank is provided based on the elapsed time.

The system proposed in this paper, does not involve heavy algorithm or extensive iterations, making it computationally efficient to be deployed on the mobile devices using native features without requiring it to process the image or analyse the features on the server (Table 6, Fig. 2(b) and Solmaz et al. (2018)). Therefore, the system is real-time and more secure. For the other R-criteria, i.e. robustness, we have also included different orientation of the sample to vary the camera to sample position. The randomness of the light source '5' in Fig. S1.2 (Supplementary Document 1) created variation in the illumination condition.

Based on the elapsed time, our work showed similar rapidness as Mutlu et al. (2017). The combined use of three different classifiers along with a larger feature-set by H. Kim et al. (2017) resulted in considerably larger model size and higher computational time.

The overall performance of the system, evaluated by the ASSURED criteria, is summarised below.

- Due to intelligent histogram based- image processing technique, the system is user-friendly. Unlike the literature (Table 8), the system is completely automatic and does not require any user intervention. To provide this autonomy without compromising the accuracy (98.94% accuracy to separate the ROI), the hybrid algorithms were

JOURNAL PRE-PROOF

optimised while developing the image processing framework in Fig. 5; the accuracy was well maintained despite the change of the condition as demonstrated using the extended dataset.

- The developed system can automatically process the image of the assays without any additional hardware component. The system is equipment-free, does not have any operating cost and accessible. Conceptually, the proposed system is deployable on the mobile platform as described in (Shabut et al., 2018) and to similar colour-based applications such as (El-Bendary et al., 2015; Smith et al., 2014).
- The system possesses the adaptability to work on similar assays without compromising the performance, confirmed by experiments conducted on the assay from a different brand (א_pH dataset) as well as urine dipstick. To the best of the authors' knowledge, there is no such evaluation for robustness performed in the literature.
- Due to less iterative image processing, optimised feature-set and selection of the classifier, the computational complexity was optimised (Fig. 10 and Table 7). The result can be produced in real-time, conceptually faster than the mentioned works in Table 8.
- The system was trained under the semi-controlled ambient condition on a balanced dataset using cross-validation. The performance was validated on completely unseen data. The system showed high accuracy, specificity and sensitivity for colour classification without compromising the degree of freedom.

9. Conclusion

This paper presented a computational system for paper-based lateral flow assays suitable to act as a standalone system on the POC platform, whether integrated to a server or not. Due to technical and economic feasibility, we have utilised universal pH indicator papers, possessing multi-objects/sample to demonstrate the proof of concept. This paper investigated, designed and developed an immaculate image processing framework to separate multiple colour pads in the universal pH indicator paper with >98% accuracy, tested and validated by varying assays. The intelligent decision-making component of the proposed frameworks, allows a complete separation of the ROI, despite the ambient condition.

After separating the ROI, this paper proposed an exclusive feature-set, i.e. pseudo-control colours to be part of the feature-set. Exploring 440 features for the LFA dataset, the optimised feature-set was found to be mean, energy and pseudo-control colours. For the stated case study, the 10-fold cross-validated training and testing for 520 samples was conducted within 0.11 seconds with 99.29-100% accuracy (95% CI). The extensive analysis based on the evaluation criteria suggested our system to be more compatible with the ASSURED criteria than existing similar works. Among the reported articles, the research conducted by Mutlu et

JOURNAL PRE-PROOF

al. (2017) is the most similar research performed in this paper. Therefore, we have deployed Mutlu's method using our dataset to present a fair comparison. Both of these studies attained 100% accuracy to provide semi-quantitative colourimetric decisions. Due to the meticulously tested reliability of the system on an adequate amount of appropriate data, our claim is well supported by the precise experimental results. The critical assessment conducted in this paper suggests our system to be more robust and more reliable due to bigger and more variant dataset and experimentations on more features and classifiers using case studies of similar applications. The reliability can be quantified from the performance on a different dataset in Fig. 19 and 95% confidence interval in

JOURNAL PRE-PROOF

Table 5.

Based on the performance of the proposed framework and insights gather from the experiments conducted, the acquired and developed knowledge regarding the image-based colourimetric system can be well utilised to similar mobile-enabled expert systems that requires visual knowledge interpretation based on the colour perception, especially where the imaging is conducted in the ambient condition (Shabut et al., 2018; El-Bendary et al., 2015).

10. Future Works

In future, more elements of robustness such as different devices and few more controlled-illumination conditions can also be considered. This should aid in rectifying shadow effect as well as improving the image segmentation accuracy more than this work (98.94%).

In addition to traditional machine learning techniques, this paper also utilised pre-trained models of deep learning. Due to the need of high processing computational systems including GPU and advanced mobile phones, deep learning including these pre-trained models deficit many attributes (A, E, D) of the ASSURED criteria at this moment. However, ensuring a better performance (S, S, R) from deep learning, its future prospect for colourimetric tests could be better, especially for telemedicine services where the system already involves a cloud-based approach.

Acknowledgement

This research is funded by the Erasmus Mundus FUSION project (Grant reference number: 2013-3254 1/001001). The authors' thank Mr Paul Cotton, Faculty of Medical Science, Anglia Ruskin University for his support during the laboratory experiments.

Declaration-of-competing-interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit Author Statement

This work is one of the case studies of the doctoral research of Dr Marzia Hoque Tania, which was supervised by Prof M. A. Hossain, Dr Khin T. Lwin and Dr Antesar M. Shabut. Prof Hossain was also responsible for the funding acquisition. Dr Mohammad Najlah provided his

JOURNAL PRE-PROOF

invaluable support to conduct the experiments in his lab, complying with health and safety laws. Dr Jeannette Chin was responsible for the critical read, which helped to authors to improve the manuscript.

JOURNAL PRE-PROOF

Appendix

Supplementary Document 1: Experimental Setup

Supplementary Document 2: Major Steps of the Image processing and Feature Extraction Algorithm

Supplementary Document 3: Performance Evaluation of the Image Processing Step

Supplementary Document 4: Statistical Analysis

References

Abuhassan, K.J., Bakhori, N.M., Kusnin, N., Azmi, U.Z.M., Tania, M.H., Evans, B.A., Yusof, N.A. and Hossain, M.A., 2017. Automatic Diagnosis of Tuberculosis Disease Based on Plasmonic ELISA and Color-based Image Classification. In: *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. [online] Jeju Island, South Korea: IEEE, pp.4512–4515. Available at: <http://ieeexplore.ieee.org/document/8037859/>.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. and Süsstrunk, S., 2012. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, [online] 34(11), pp.2274–2282. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6205760&isnumber=6304885%0A>.

Akraa, S., Anh Pham, T.T., Shen, H., Tang, Y., Tang, B.Z., Li, J. and Walker, S., 2018. A smartphone-based point-of-care quantitative urinalysis device for chronic kidney disease patients. *Journal of Network and Computer Applications*. [online] Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1084804518301449> [Accessed 1 May 2018].

Alankus, G., Horzum, N., Mutlu, A.Y., Bayram, A. and Solmaz, M.E., 2018. Single-Image-Referenced Colorimetric Water Quality Detection Using a Smartphone. *ACS Omega*, 3(5), pp.5531–5536.

Alidans srl, 2015. *AssayColor*. [online] Android App on Google Play. Available at: <https://play.google.com/store/apps/details?id=com.alidans.assaycolor> [Accessed 10 Jan. 2017].

Arthur, D. and Vassilvitskii, S., 2007. k-means++: The Advantages of Careful Seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. [online] New Orleans, Louisiana, USA, pp.1027–1035. Available at: <http://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf> [Accessed 12 Dec. 2017].

JOURNAL PRE-PROOF

Barbosa, A.I., Gehlot, P., Sidapra, K., Edwards, A.D. and Reis, N.M., 2015. Portable smartphone quantitation of prostate specific antigen (PSA) in a fluoropolymer microfluidic device. *Biosensors and Bioelectronics*, [online] 70, pp.5–14. Available at: <http://www.sciencedirect.com/science/article/pii/S0956566315001499>.

Bourouis, A., Feham, M., Hossain, M. a. and Zhang, L., 2014. An intelligent mobile based decision support system for retinal disease diagnosis. *Decision Support Systems*, [online] 59(1), pp.341–350. Available at: <http://dx.doi.org/10.1016/j.dss.2014.01.005>.

Bradley, D. and Roth, G., 2007. Adaptive Thresholding using the Integral Image. *Journal of Graphics Tools*, 12(2), pp.13–21.

Carbonera, J.L., Abel, M. and Scherer, C.M.S., 2015. Visual interpretation of events in petroleum exploration: An approach supported by well-founded ontologies. *Expert Systems With Applications*, [online] 42(5), pp.2749–2763. Available at: <http://www.sciencedirect.com/science/article/pii/S0957417414007076>.

Chen, C., Wu, Y. and Dong, T., 2014. Dipsticks integrated on smart diapers for colorimetric analysis of urinary tract infections in the field. In: *Proceedings of the 16th International Conference on Mechatronics, Mechatronika 2014*. [online] Brno, Czech Republic, pp.423–427. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7018295&tag=1>.

Colour Blind Awareness, 2018. *Colour Blindness*. [online] 2017. Available at: <http://www.colourblindawareness.org/colour-blindness/> [Accessed 23 Jun. 2018].

Contreras-naranjo, J.C., Wei, Q. and Ozcan, A., 2016. Mobile Phone-Based Microscopy , Sensing , and Diagnostics. *IEEE Journal of Selected Topics in Quantum Electronics*, [online] 22(3), pp.1–14. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7265008>.

Cooper, D., Callahan, B., Callahan, P. and Burnett, L., 2012. Mobile Image Ratiometry: A New Method for Instantaneous Analysis of Rapid Test Strips. In: *Nature Precedings*. [online] pp.2–3. Available at: <http://precedings.nature.com/documents/6827/version/1>.

Dang, L.M., Hassan, S.I., Im, S. and Moon, H., 2019. Face image manipulation detection based on a convolutional neural network. *Expert Systems With Applications*, [online] 129, pp.156–168. Available at: <http://www.sciencedirect.com/science/article/pii/S0957417419302350>.

Dhar, A., Mehta, R. and Sit, M., 2017. *Colorimetric Detection of pH Strips*. Available at:

JOURNAL PRE-PROOF

<https://github.com/mattsit/ph-machine-learning>.

El-Bendary, N., El Hariri, E., Hassanien, A.E. and Badr, A., 2015. Using machine learning techniques for evaluating tomato ripeness. *Expert Systems with Applications*, [online] 42(4), pp.1892–1905. Available at: <https://www.sciencedirect.com/science/article/pii/S0957417414006186> [Accessed 10 Jun. 2019].

Enzo Life Sciences inc., 2015. *Enzo ELISA Plate Reader*. [online] Android App on Google Play. Available at: <https://play.google.com/store/apps/details?id=com.enzo.elisaplatereader> [Accessed 21 Sep. 2017].

Feng, S., Caire, R., Cortazar, B., Turan, M., Wong, A. and Ozcan, A., 2014. *Immunochromatographic diagnostic test analysis using google glass*. *ACS Nano*, .

Garg, S., Ramprasaath, R.S., Kapur, S. and Rao, K.M.M., 2014. Automated colorimetric analysis in paper based sensors. In: *Image Processing (ICIP), 2014 IEEE International Conference on*. [online] Paris, pp.3607–3611. Available at: <http://ieeexplore.ieee.org/document/7025732/>.

GSMA Intelligence, 2017. *Definitive data and analysis for the mobile industry*. [online] Available at: <https://www.gsmaintelligence.com/> [Accessed 7 Aug. 2017].

He, K., Zhang, X., Ren, S. and Sun, J., 2015. Deep Residual Learning for Image Recognition. [online] Available at: <https://arxiv.org/pdf/1512.03385.pdf> [Accessed 10 Jun. 2018].

Hussain, I., Das, M., Ahamad, K.U. and Nath, P., 2017. Water salinity detection using a smartphone. *Sensors and Actuators, B: Chemical*, [online] 239, pp.1042–1050. Available at: <http://dx.doi.org/10.1016/j.snb.2016.08.102>.

Janke, J., Castelli, M. and Popovič, A., 2019. Analysis of the proficiency of fully connected neural networks in the process of classifying digital images: Benchmark of different classification algorithms on high-level image features from convolutional layers. *Expert Systems With Applications*, [online] 135, pp.12–38. Available at: <http://www.sciencedirect.com/science/article/pii/S0957417419303938>.

Jonas, S.M., Deserno, T.M., Buhimschi, C.S., Makin, J., Choma, M.A. and Buhimschi, I.A., 2016. Smartphone-based diagnostic for preeclampsia: An mHealth solution for administering the Congo Red Dot (CRD) test in settings with limited resources. *Journal of the American Medical Informatics Association*, [online] 23(1), pp.166–173. Available at:

JOURNAL PRE-PROOF

<https://academic.oup.com/jamia/article-lookup/doi/10.1093/jamia/ocv015>.

Karlsen, H., 2018. *Smartphone-based urinary biomarker detection: an application-oriented device and algorithm*. [online] University College of Southeast Norway. Available at: <https://brage.bibsys.no/xmlui/handle/11250/2480144> [Accessed 21 Mar. 2018].

Karlsen, H. and Dong, T., 2017. Smartphone-Based Rapid Screening of Urinary Biomarkers. *IEEE Transactions on Biomedical Circuits and Systems*, [online] 11(2), pp.455–463. Available at: <http://ieeexplore.ieee.org/document/7879269/> [Accessed 16 Mar. 2017].

Kettler, H., White, K. and Hawkes, S., 2004. *Mapping the landscape of diagnostics for sexually transmitted infections: Key Findings and Recommendations*. [online] Available at: <http://www.who.int/tdr/publications/documents/mapping-landscape-sti.pdf?ua=1> [Accessed 23 Apr. 2018].

Khademhosseini, A., 2011. Nano/microfluidics for diagnosis of infectious diseases in developing countries. *Adv Drug Delivery Rev*, [online] 62(4–5), pp.449–457. Available at: <http://dx.doi.org/10.1016/j.addr.2009.11.016>.

Khan, M.S. and Garnier, G., 2013. Direct measurement of alkaline phosphatase kinetics on bioactive paper. *Chemical Engineering Science*, [online] 87, pp.91–99. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0009250912005891> [Accessed 15 Jul. 2017].

Kim, H., Awofeso, O., Choi, S., Jung, Y. and Bae, E., 2017a. Colorimetric analysis of saliva--alcohol test strips by smartphone-based instruments using machine-learning algorithms. *Appl. Opt.*, [online] 56(1), pp.84–92. Available at: <http://ao.osa.org/abstract.cfm?URI=ao-56-1-84>.

Kim, S.C., Jalal, U.M., Im, S.B., Ko, S. and Shim, J.S., 2017b. A smartphone-based optical platform for colorimetric analysis of microfluidic device. *Sensors and Actuators, B: Chemical*, [online] 239, pp.52–59. Available at: <http://dx.doi.org/10.1016/j.snb.2016.07.159>.

Koczula, K.M. and Gallotta, A., 2016. Lateral flow assays. *Essays in biochemistry*, [online] 60(1), pp.111–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27365041> [Accessed 20 Jun. 2018].

Konnaiyan, K., Cheemalapati, S., Pyayt, A. and Gubanov, M., 2017. mHealth Dipstick Analyzer For Monitoring of Pregnancy Complications. *IEEE Sensors Journal*, 17(22), pp.7311–7316.

JOURNAL PRE-PROOF

Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in neural information processing systems*. [online] Neural Information Processing Systems Foundation, Inc., pp.1097–1105. Available at: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> [Accessed 9 Apr. 2018].

Lopez-Ruiz, N., Curto, V.F., Erenas, M.M., Benito-Lopez, F., Diamond, D., Palma, A.J. and Capitan-Vallvey, L.F., 2014. Smartphone-based simultaneous pH and nitrite colorimetric determination for paper microfluidic devices. *Analytical Chemistry*, [online] 86(19), pp.9554–9562. Available at: <http://pubs.acs.org/doi/abs/10.1021/ac5019205?journalCode=ancham>.

Masawat, P., Harfield, A. and Namwong, A., 2015. An iPhone-based digital image colorimeter for detecting tetracycline in milk. *Food Chemistry*, [online] 184, pp.23–29. Available at: <http://dx.doi.org/10.1016/j.foodchem.2015.03.089>.

Mutlu, A.Y., Kılıç, V., Özdemir, G.K., Bayram, A., Horzum, N. and Solmaz, M.E., 2017. Smartphone-based colorimetric detection via machine learning. *The Analyst*, [online] 142(13), pp.2434–2441. Available at: <http://xlink.rsc.org/?DOI=C7AN00741H>.

NHS Choices, 2016. *Colour vision deficiency (colour blindness)*. [online] NHS.UK. Available at: <https://www.nhs.uk/conditions/colour-vision-deficiency/> [Accessed 25 Jul. 2018].

Otsu, N., 1979. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, [online] 9(1), pp.62–66. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4310076&isnumber=4310064>.

Ozkan, H., 2017. Rapid Daignostic Lateral Flow Strip Test Reader. *BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING*, 5(2), pp.34–39.

P. D. of the Department of Economic and S. A. of the United Nations Secretariat, 2012. *World population prospects: The 2012 revision*. [online] UN. Available at: <http://www.un.org/en/development/desa/population/publications/pdf/trends/WPP2012_Wallchart.pdf> [Accessed 27 Jun. 2017].

Rahmat, R.F., Royananda, Muchtar, M.A., Taqiuddin, R., Adnan, S., Anugrahwaty, R. and Budiarto, R., 2018. Automated color classification of urine dipstick image in urine examination. *Journal of Physics: Conference Series*, [online] 978(1), p.012008. Available at: <http://stacks.iop.org/1742-6596/978/i=1/a=012008?key=crossref.88f0a381b969099cf5bf5e6b9c95e157> [Accessed 20

JOURNAL PRE-PROOF

Mar. 2018].

Rajan, A. and Glorikian, H., 2009. Point-of-care diagnostics: market trends and growth drivers. *Expert Opinion on Medical Diagnostics*, [online] 3(1), pp.1–4. Available at: <http://www.tandfonline.com/doi/abs/10.1517/17530050802651579?journalCode=iedg20> [Accessed 3 Nov. 2015].

Rasmussen, C.E., Williams, C.K.I., Sutton, R.S., Barto, A.G., Spirtes, P., Glymour, C., Scheines, R., Schölkopf, B. and Smola, A.J., 2006. Gaussian Processes for Machine Learning. In: T. Dietterich, ed., *Adaptive Computation and Machine Learning*. [online] MIT Press . Available at: <http://www.gaussianprocess.org/gpml/chapters/RW.pdf> [Accessed 9 Jun. 2018].

Roda, A., Michelini, E., Zangheri, M., Di Fusco, M., Calabria, D. and Simoni, P., 2016. Smartphone-based biosensors: A critical review and perspectives. *TrAC - Trends in Analytical Chemistry*, [online] 79, pp.317–325. Available at: <http://dx.doi.org/10.1016/j.trac.2015.10.019>.

Seo, Y. and Shin, K., 2019. Hierarchical convolutional neural networks for fashion image classification. *Expert Systems With Applications*, [online] 116, pp.328–339. Available at: <http://www.sciencedirect.com/science/article/pii/S0957417418305992>.

Sergyan, S., 2008. Color histogram features based image classification in content-based image retrieval systems. *2008 6th International Symposium on Applied Machine Intelligence and Informatics*, [online] pp.221–224. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4469170&tag=1>.

Shabut, A.M., Hoque Tania, M., Lwin, K.T., Evans, B.A., Yusof, N.A., Abu-Hassan, K.J. and Hossain, M.A., 2018. An intelligent mobile-enabled expert system for tuberculosis disease diagnosis in real time. *Expert Systems with Applications*, [online] 114, pp.65–77. Available at: <https://www.sciencedirect.com/science/article/pii/S0957417418304214> [Accessed 21 Jul. 2018].

Shen, L., Hagen, J. a. and Papautsky, I., 2012. Point-of-care colorimetric detection with a smartphone. *Lab on a Chip*, [online] 12(21), p.4240. Available at: <http://pubs.rsc.org/en/content/articlelanding/2012/lc/c2lc40741h/unauth#!divAbstract>.

Sicard, C., Glen, C., Aubie, B., Wallace, D., Jahanshahi-Anbuhi, S., Pennings, K., Daigger, G.T., Pelton, R., Brennan, J.D. and Filipe, C.D.M., 2015. Tools for water quality monitoring and mapping using paper-based sensors and cell phones. *Water Research*, [online] 70,

JOURNAL PRE-PROOF

pp.360–369. Available at: <http://ac.els-cdn.com/S0043135414008379/1-s2.0-S0043135414008379-main.pdf?_tid=939ea938-be36-11e6-9371-00000aacb361&acdnat=1481305409_ac406ff286a0ae93cf2692fa0fedef3f>.

Sicasys Software GmbH, 2017. *Spotxel® Reader*. [online] Google Play. Available at: <https://play.google.com/store/apps/details?id=com.sicasys.spotxel&hl=en> [Accessed 12 Jan. 2018].

Siemens Healthcare GmbH, 2018. *CLINITEK Status+ Analyzer*. [online] Available at: <https://www.healthcare.siemens.com/urinalysis-products/urinalysis-systems/clinitek-status-analyzer> [Accessed 19 May 2018].

Smith, G.T., Dwork, N., Khan, S.A., Millet, M., Magar, K., Javanmard, M., Ellerbee Bowden, A.K., Patel, H., Livsey, S., Swann, R., Bukhari, S., Devillé, W.L., Yzermans, J.C., Duijn, N.P. Van, Bezemer, P.D., Windt, D.A. Van Der, Bouter, L.M., Abirami, K., Tiwari, S., Yetisen, A.K., Martinez-Hurtado, J., Garcia-Melendrez, A., Vasconcellos, F. da C., Lowe, C.R., Mudanyali, O., Dimitrov, S., Sikora, U., Padmanabhan, S., Navruz, I., Ozcan, A., Martinez, A.W., Phillips, S.T., Carrilho, E., III, S.W.T., Sindi, H., Whitesides, G.M., Martinez, A.W., Phillips, S.T., Whitesides, G.M., Du, W., Li, L., Nichols, K.P., Ismagilov, R.F., Shen, F., Davydova, E.K., Du, W., Kreutz, J.E., Piepenburg, O., Ismagilov, R.F., Shen, F., Du, W., Davydova, E.K., Karymov, M.A., Pandey, J., Ismagilov, R.F., Liu, W., Chen, D., Du, W., Nichols, K.P., Ismagilov, R.F., Hong, J.I., Chang, B.-Y., Shen, L., Hagen, J.A. and Papautsky, I., 2016. Robust dipstick urinalysis using a low-cost, micro-volume slipping manifold and mobile phone platform. *Lab Chip*, [online] 58, pp.951–954. Available at: <http://xlink.rsc.org/?DOI=C6LC00340K>.

Smith, J.E., Griffin, D.K., Leny, J.K., Hagen, J. a, Chávez, J.L. and Kelley-Loughnane, N., 2014. Colorimetric detection with aptamer-gold nanoparticle conjugates coupled to an android-based color analysis application for use in the field. *Talanta*, [online] 121, pp.247–55. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24607135>.

Solmaz, M.E., Mutlu, A.Y., Alankus, G., Kılıç, V., Bayram, A. and Horzum, N., 2018. Quantifying colorimetric tests using a smartphone app based on machine learning classifiers. *Sensors and Actuators B: Chemical*, [online] 255, pp.1967–1973. Available at: <https://www.sciencedirect.com/science/article/pii/S0925400517316519> [Accessed 24 Feb. 2018].

Statista, 2015. *Number of mobile phone users worldwide 2013-2019*. [online] The Statistics Portal. Available at: <https://www.statista.com/statistics/274774/forecast-of-mobile-phone-

JOURNAL PRE-PROOF

users-worldwide/> [Accessed 24 Apr. 2018].

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B. and Vandewalle, J., 2002. *Least Squares Support Vector Machines*. [online] WORLD SCIENTIFIC. Available at: <https://www.worldscientific.com/worldscibooks/10.1142/5089> [Accessed 6 Aug. 2018].

Suykens, J.A.K. and Vandewalle, J., 1999. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, [online] 9(3), pp.293–300. Available at: <http://link.springer.com/10.1023/A:1018628609742> [Accessed 6 Aug. 2018].

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [online] IEEE, pp.1–9. Available at: <http://ieeexplore.ieee.org/document/7298594/> [Accessed 8 May 2018].

Tania, M.H., Lwin, K.T., Abuhassan, K. and Bakhori, N.M., 2017. An Automated Colourimetric Test by Computational Chromaticity Analysis: A Case Study of Tuberculosis Test. In: *Advances in Intelligent Systems and Computing*. [online] Springer, Cham, pp.313–320. Available at: <http://link.springer.com/10.1007/978-3-319-60816-7>.

Tania, M.H., Lwin, K.T. and Hossain, M.A., 2016. Computational complexity of image processing algorithms for an intelligent mobile enabled tongue diagnosis scheme. In: *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*. [online] Chengdu: IEEE, pp.29–36. Available at: <http://ieeexplore.ieee.org/document/7916193/>.

Tania, M.H., Lwin, K.T., Shabut, A.M. and Hossain, M.A., 2018. Clustering and Classification of a Qualitative Colorimetric Test. In: *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. [online] Southend, United Kingdom: IEEE, pp.7–11. Available at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8658480&isnumber=8658427>.

Vashist, S.K., van Oordt, T., Schneider, E.M., Zengerle, R., von Stetten, F. and Luong, J.H.T., 2015. A smartphone-based colorimetric reader for bioanalytical applications using the screen-based bottom illumination provided by gadgets. *Biosensors and Bioelectronics*, [online] 67, pp.248–255. Available at: <http://dx.doi.org/10.1016/j.bios.2014.08.027>.

Wang, S., Xu, F. and Demirci, U., 2010. Advances in developing HIV-1 viral load assays for resource-limited settings. *Biotechnology Advances*, [online] 28(6), pp.770–781. Available at: <http://dx.doi.org/10.1016/j.biotechadv.2010.06.004>.

JOURNAL PRE-PROOF

Wang, Y., Li, Y., Bao, X., Han, J., Xia, J., Tian, X. and Ni, L., 2016. A smartphone-based colorimetric reader coupled with a remote server for rapid on-site catechols analysis. *Talanta*, [online] 160, pp.194–204. Available at: <http://dx.doi.org/10.1016/j.talanta.2016.07.012>.

Wirth, M., Biswas, N., Ahmad, S., Nayak, H.S., Pugh, A., Gupta, T. and Mahmood, I., 2018. A prospective observational pilot study to test the feasibility of a smartphone enabled uChek© urinalysis device to detect biomarkers in urine indicative of preeclampsia/eclampsia. *Health and Technology*, [online] pp.1–6. Available at: <https://doi.org/10.1007/s12553-018-0248-0> [Accessed 11 Aug. 2018].

World Bank Group, 2016. *World Development Report 2016: Digital Dividends*. [online] Washington DC. Available at: <http://elibrary.worldbank.org/doi/book/10.1596/978-1-4648-0671-1> [Accessed 10 May 2018].

World Health Organization, 2017. *Global Health Observatory (GHO) data*. [online] WHO. Available at: <http://www.who.int/gho/health_workforce/physicians_density/en/> [Accessed 27 Jun. 2017].

Yetisen, A.K., Martinez-Hurtado, J.L., Garcia-Melendrez, A., da Cruz Vasconcellos, F. and Lowe, C.R., 2014. A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests. *Sensors and Actuators B: Chemical*, [online] 196, pp.156–160. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S092540051400094X>.

JOURNAL PRE-PROOF